AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 115. SEPTEMBER, 1916.

SCOPE AND METHOD OF STATISTICS.

By Harald Westergaard, *Professor of Social Science in the University of Copenhagen.*

INTRODUCTION.

By Walter F. Willcox.

It is a pleasure to welcome to the pages of these *Publications* the writer of the following weighty article and to express my own debt to him, a debt which began almost a quarter of a century ago with the beginning of my work as a teacher of statistics and has steadily deepened with the passage of the years. I know of hardly any writer in this field from whom I have learned more and whose work I can commend to others more unreservedly. That I am not alone among Americans in this opinion is suggested by the fact that years ago Professor Westergaard was invited by the United States to represent the guild of European statists at the Congress of Arts and Sciences in St. Louis and to give there an address upon demography. As he was unable then to accept, the present article may be deemed his first appearance before an American public and it seems appropriate, therefore, for me to add a few personal words about his career.

The son of a distinguished Oriental scholar, he himself turned first to mathematics and later to the social sciences. In these departments his academic work was supplemented by extended journeys in Germany, Austria, Switzerland, France, and England, during which he studied social problems, statistics, and insurance. At the age of 28 he presented to the University of Copenhagen a work on mortality outlining

and appraising recent investigations of the death rate in various classes of the population. This brought to him the University's gold medal offered as a prize for the best treatise on that topic. His work was published in German under the title, *Principles of Mortality and Morbidity*, and at the same time the author, who had already been employed for two years in the Danish Insurance Office, became a professor of social science at the University which had thus crowned his work. His book is an effort to analyze the various causes or influences, such as age, sex, marital condition, environment (urban or rural), race, religion, occupation, and diet, which affect man's death rate and to measure, or at least to point out the sound method for measuring, the effect of each influence upon mortality. Nearly twenty years later it appeared in a new edition completely rewritten and greatly enriched by the results of the author's intervening studies and those of other scholars in the same subject. For years I have found it an indispensable repository of fact and above all of argument, critical and constructive, in the wide field it covers.

Professor Westergaard's second book, appearing after an interval of eight years, grew directly out of his teaching. It is an *Outline of the Theory of Statistics*, published simultaneously in Danish and German editions and designed especially for the use of students in social science at the University of Copenhagen. Perhaps no better introduction to the present article can be found than a few words from the preface to his second main work:

Writers on the Theory of Statistics are often satisfied with developing the methods employed in special fields, like mortality, merely adding a few general remarks about the regularity of statistical phenomena and its consistency with our ideas of human freedom. This blinks the main question; namely, Under what conditions and within what limits is the regularity of statistical phenomena, "the law of large numbers," manifested? To sidestep that question has advantages, because it relieves one from entering upon a mathematical analysis; but it has also serious disadvantages, because it deprives statistics of its foundation. In following that procedure the statist cannot decide whether he has a solid basis

for his work because he does not know whether the law of large numbers is exemplified in the material before him. In the first part of the following book I have set forth the fundamental principles of the calculus of probabilities and then undertaken a task often overlooked, that of investigating whether these principles are exemplified in the results of statistical observation.

This second work of Westergaard, like the first, has been thoroughly revised and rewritten within the quarter century since it was published and is now about to appear in Danish. I sincerely hope that it may appear also in an English edition.

The present article is an epitome of the author's mature conclusions on a subject upon which American statists need to profit from the results of European scholars. The author's familiarity with the best English experience, perhaps suggested by the fact that the first edition of each of his two main works is dedicated to an English scholar and insurance expert, appears on nearly every page. He is wonderfully well informed also regarding the development of his subject from the beginning and his chapters on the history of statistics and of mortality statistics are marked by a combination of erudition and analytical power which raises them high above almost any other work and makes them a model in that field. For American scholars this is the more important because hardly one even of our scholars' libraries possesses the books needed for independent investigations in the history of statistics and yet perhaps no study provides a better thread to guide our researches past the pitfalls which trapped many former investigators and into the fruitful fields of statistical analysis.

In another respect, also, Professor Westergaard has much to teach our American statists. Those persons, mainly government officials, who are engaged in prosecuting statistical inquiries are seldom inclined by interest and training to use subtle or unusual methods of refined analysis upon the figures they have produced and thus fail to extract their full meaning. But they often have a just and sometimes an exaggerated appreciation, developed by their own sad experience, of the many errors from which their results suffer and of the uncertainty or at least the considerable margin of error in their

conclusions. This seems to be true of nearly every official American statist since Walker and Wright. On the other hand, such American scholars, trained in higher mathematics, as apply refined and modern methods to statistical data are, in the main, university men whose experience has seldom given them an opportunity to estimate the margin of error in official returns. They frequently accept without question figures so wide of the truth as to be most unsafe. For example, my own experience, including membership on a committee appointed to probe the divergencies between the figures of the Census Office and those of the Department of Agriculture, has made me sceptical of the annual returns showing acreage and yield of the staple American crops, returns now published under the more descriptive title of crop estimates. For this reason I have not thought it safe to use them in any published discussions. And yet elaborate theoretical structures employing the most refined methods have been built on these figures apparently without inquiry as to whether the foundation is strong enough to support the superstructure. These two classes of investigators, the producers of statistical data and the analysts of the results, need a closer coöperation and a more intelligent sympathy with each other's work.

It is noteworthy that Professor Westergaard, whose contributions to the refinements of statistical method have been of the first importance, should close his present article with the suggestion that our first object "must be not so much to prepare refined statistical methods as to provide useful observations. . . . At present we are more in need of statistical data than of theoretical investigations. . . . We want statistical observations covering new fields and here an enormous amount of work remains to be done" (page 52). In the United States both types of work are sorely needed, but more important than either alone is a sympathetic comprehension by each group of the aims and the difficulties of the other. In this way, and in this way alone, the new material needed can be most rapidly secured and the new methods applied only to significant material. Such a lesson of open-eyed and sympathetic coöperation Professor Westergaard is preëminently qualified to teach.

SCOPE AND METHOD OF STATISTICS.

By Harald Westergaard.

I.

It might appear absurd to discuss the scope and method of a science with at least 250 years of history, an evolution from the trifling beginning as "political arithmetic" to its present position as an important, or rather an indispensable, instrument in the complicated machinery of modern government and an auxiliary science in many branches of human thought. It might be supposed that no unsettled questions about scope and method were left, that on the contrary the tradition of many years would enable the great army of statisticians to follow a definite plan, without any doubt whatever about their final goal.

Still it is a fact, that there are at present not one, but several corps of statisticians, each trying earnestly to promote the science, but hardly able to coöperate for lack of mutual sympathy and sometimes acting in direct opposition to one another.

I hope, therefore, that it will not appear entirely useless to try to throw some little light on this problem. In doing so I shall first give a brief outline of some essential points in the evolution of statistics.

In the seventeenth and eighteenth centuries we find three entirely different lines of evolution, viz., the calculus of probabilities, "political arithmetic," and the comparative description of states, the last discipline being the only one to which the name of statistics was appropriated, though this one had originally very little to do with statistics in the modern sense of the word.

Statistics as the comparative description of states was cultivated by Aristotle in his famous Constitutions of 158 states, of which only the Constitution of Athens has survived. To the same category belong the works of the Italian authors Sansovino and Botero, dating from the end of the sixteenth century, those of their contemporary in France, Etienne Pasquier, who wrote *Récherches de la France*, and those of Pierre d'Avity, who published his great work, *Les Estats du Monde*, a little later (1614).

It was, however, in Germany chiefly that this study was cultivated. The well-known polyhistor, Hermann Conring, lectured on it for many years in the seventeenth century, beginning his description with Spain and ending it with Japan, Morocco, and Abyssinia; and in the eighteenth century we meet in Germany a whole series of writers from Achenwall to Schlözer, dealing with this theme. Most of these works had a family resemblance containing descriptions of the several countries, their climate and population, their produce and constitution, etc., but with hardly any numerical data; for these German authors did not take much interest even in the scanty numerical material which was at hand. Sometimes these descriptions were printed in a schematic form, giving under each heading for each country a short verbal description; in this way the so-called "Tabellen-Statistik" arose.

These tabular descriptions might be looked upon as signs of degeneration; but, clumsy and naïve as they often were, they actually indicate progress. For gradually the authors of *Tabellen-Statistik* began to fill their paragraphs with numerical facts, as material of this kind accumulated. In particular comparative tables with facts relating to population were introduced. One of the most important contributions of this kind was written by Crome (1785), with tables containing the number of inhabitants, area, density of population, and the "possible number," that is to say, the population assuming an average density of 3,000 inhabitants per (German) square mile. Crome tried to criticize in a really scientific spirit all the figures at hand and to make the best possible estimates. Thus a bridge was built between this discipline and statistics in the modern sense of the word, these comparative descriptions of various countries being increasingly influenced by the growing abundance of numerical material. This process was not without its dramatic features. At the beginning of the nineteenth century an interesting discussion over the meaning of statistics took place in some learned German publications; the views of the antagonists seemed quite irreconcilable; each side maintained that its conception of statistics was the true one. It was maintained that the adherents of the *Tabellen-Statistik* simply "skeletonized" the noble science of statistics, that they were *Tabellenknechte,*

unimaginative "slaves of the tables." As in so many other scientific controversies, there was never a formal conclusion; by and by the discussion died away, and gradually the name *statistics* was transferred from this discipline to political arithmetic, as formally proposed in a masterly essay, written in the middle of the century by the famous German economist, K. Knies.

The foundation of the calculus of probabilities, the second line of evolution, was laid by Italian and French mathematicians, who were interested in the problems of games at dice, lotteries, etc. This discipline reached a very high development in the eighteenth century, culminating at the beginning of the nineteenth century in Laplace's masterly *Théorie analytique des probabilités* (1812).

The chief problem of the calculus of probabilities was discussed in J. Bernoulli's posthumous work, *Ars conjectandi* (1713) and further developed by several brilliant mathematicians. According to Bernoulli's theorem, the probability of a certain deviation from a given type will depend on the number of observations, so that by increasing the number of observations it is possible to get as narrow limits of deviation as we like. This is the famous "law of large numbers," which has had so prominent a part in many statistical discussions. In fact, if it is possible to calculate the limits of deviation from the standard, we can easily control statistical conclusions. If the average rate of infant mortality is, say, 0.20, and if in a group of 10,000 infants a rate of 0.21 is very rarely reached, in a group of 1,000,000 infants a rate of 0.201 will be equally seldom found and if there are 100 millions of observations, the limit is reduced to 0.2001.

Although the mathematicians clearly understood the significance of Bernoulli's theorem, they failed to exhaust it fully. Most of the mathematicians who developed the calculus of probabilities seemed to suppose, as an axiom which needed no test, that its theorems would hold good everywhere. Those theorems can in fact be applied to a goodly number of cases, not the least in games at dice, lotteries, etc; but the important question was left untouched, whether and under what conditions these theorems might be applied to the more complicated problems in vital or economic statistics. No in-

vestigation was made to test the applicability of the theorems in this field; hardly even was an attempt made with regard to the original field of the calculus of probabilities, in spite of the severe but rather unjust criticism by d'Alembert, who directly appealed to mathematicians to make experiments with regard to games in order to find the real laws of chance. Many questions of this kind still remain unsolved.

The third line of evolution began in England. The cradle of political arithmetic is to be found in London, where a merchant, John Graunt, in 1662 wrote his remarkable *Natural and Political Observations upon the Bills of Mortality.* This treatise is a keen and ingenious attempt to calculate a life table from the London bills of mortality, on the basis of the returns of cause of death only, no statement of age at death being made. Graunt's results probably were wide of the truth; thus he supposed that three eighths of the persons surviving at six years would die before reaching sixteen, and again, that the same proportion would die between sixteen and twenty-six. Probably the actual death rate was only a fraction of this estimate. But still we must admire Graunt's acumen and his grasp of these highly important problems, which in the following generations were to be studied by a number of keen investigators, including the brilliant Halley and Lavoisier.

Halley's Breslau table is the next step in advance, a real seven-league stride, hardly appreciated by his contemporaries. His paper, published in 1693, contained *in nuce* a whole theory of vital statistics and life insurance. And not only did he grasp the problems thoroughly, but there is also reason to believe that his numerical results were fairly correct, in spite of the imperfections of his data. Thus, the total number of inhabitants which he calculated seems to have been fairly close to the truth, and so, too, his life table; it gave a good picture of the chances of death in a non-epidemic period in those times, though mortality in years of plague might, indeed, rise to a much higher level.

In the following century several mortality tables were constructed, often based on much more complete observations than were at Halley's disposal; thus by Struyck and Kersse-

boom in Holland and Deparcieux in France, the latter for instance calculating most interesting life tables for monks and nuns. Not less important was the progress in Sweden (1748), where a whole system of vital statistics was founded, with a population register in each parish and complete lists of births and deaths. The first reliable mortality tables for a whole country were calculated by Wargentin from these data.

How rich the literature of political arithmetic had grown in the course of a century appears from Süssmilch's compilation in his great work, *Die göttliche Ordnung* (2nd ed., 1761). His work contained, moreover, several original contributions, including a life table which enjoyed a great reputation in spite of conspicuous defects, obvious even to his contemporaries.

Halley's contemporaries, King and Davenant, had chiefly economic statistics in view. We owe to Davenant a first attempt at a theory of political arithmetic. (*Of the Use of Political Arithmetic, in all considerations about the revenues and trade*, 1698.) A single quotation will suffice: "As for example, when the number of inhabitants in England is known, by considering the extent of the French territory, their way of living, and their soil, and by comparing both places, and by other circumstances, a near guess may be made how many people France may probably contain." Statistical conclusions were very often drawn in this way in those days. Unfortunately the keystone was wanting; Davenant failed to explain how to avoid erroneous conclusions or how to find the limits of probable error.

In France several renowned authors of the eighteenth century grappled with the problem of calculating the number of inhabitants. So Messance, who, in his *Récherches sur la population* (1766), found a population of 24,000,000; Moheau arrived at a similar result (1778). Their method is easily understood. Moheau finds, for instance, in representative districts, that each year on an average there would be one birth to every 25½ inhabitants; the average number of births in France being about 929,000, he estimates the population at 23,700,000. The number of deaths are in Moheau's opinion less reliable as a basis for estimate, though in the

present case, by using the deaths, he was able to reach nearly the same result.

Lavoisier made a step forward in political arithmetic (1791); taking Moheau's method of estimate from representative districts, he applied it to agricultural statistics, and calculated on the basis of representative observations the number of ploughs and of horses and cattle, as well as the harvest and the cultivated area of the country.

Ingenious as these calculations often were, they were somewhat unreliable and arbitrary, nevertheless, for it seemed hardly possible to check the results. Laplace pointed in the right direction, basing his calculations of the number of inhabitants on figures of population and births in selected districts in 30 departments. His estimate of the number of inhabitants was probably nearer the truth than the previous estimates. But even he did not reach the bottom of the problem; he did not prove that these observations were grouped around a typical value according to the calculus of probabilities. After Laplace not much was done to develop the theory of representative statistics or of indirect methods of finding the numbers required. This is probably connected with the fact that direct methods of complete enumeration, such as a census covering an entire country, gradually became popular in spite of the great distrust with which at first they were regarded. In the earlier period of statistics this distrust was very characteristic. Davenant preferred the books of the Hearth Office; Messance would even rely on statistics of the yearly consumption of goods as a basis for estimating the number of inhabitants. Even as late as the middle of the nineteenth century, there was in France and Belgium a lively discussion of the advisability of using census results in calculating mortality tables, several well-known statisticians trying to do without the census.

The census held its ground, however, and no modern statistician would doubt the general reliability of enumerations of this kind. Only in calculating the infant mortality most statisticians still prefer the births to the census results, but even here a tendency toward using census figures is perceptible.

At the beginning of the nineteenth century the outlook for statistics was rather favorable. Statistical offices were established in France and Prussia, a census was taken in the United States in 1790 and again ten years later. But in Europe the long war and the political reaction after the conclusion of peace delayed progress, so that the new era could not begin until about 1830. From that time on there has been continuous and rapid progress. The enthusiasm of those early years was astonishing: statistical societies were founded; journals like *Annales d'hygiène publique* or the *Journal of the Statistical Society* of London published numerous statistical articles; statistical offices were founded or reorganized. This enthusiasm has waned a little, but nothing has prevented a steady progress, an immense accumulation of numerical data, of reports, *enquêtes*, etc. Under this rapid evolution statistics could not be altogether free from dilettanteism; it is even surprising that the number of such untrained workers has been so small.

Quetelet (1796–1874) may be considered as the typical statistician of his period. His enthusiasm, his readiness to draw conclusions from a rather meagre collection of material, his faith in statistics as the queen among all sciences, his theory of the great regularity in all statistical phenomena—all these features were common to many statisticians of his time, though they lacked the impress of his brilliant genius and his wonderful style. It is easy now to see his defects: Quetelet was too willing to believe in the constancy of types; he was fully convinced, for instance, that mortality did not change, though, even while he was writing, a considerable decrease of mortality was beginning. He saw too much in his "*homme moyen*," that remarkable being, who should unite all the qualities of the nation, as a representative of beauty and harmony; he was too optimistic with regard to international statistics and the usefulness of congresses. All this prevented him from seeing the variety which prevails everywhere, the constant changes in all conditions of human life. These mistakes were not peculiar to him. Mathematicians sought to find a formula by which mortality might be expressed as a function of age, a physiological law of mortality, so to speak.

The regularity of all statistical data being an axiom, it did not occur to the statisticians of the period that a thorough investigation into the conditions and the limits of this regularity was necessary.

But these erroneous ideas have to a great extent disappeared. In the following generation we perceive a certain reaction, a suspicion of laws and types. The statisticians mostly confined themselves to collecting exact details, giving, so to speak, a photograph of the conditions, but taking less interest in the causes behind all the phenomena. The doctrines of the present Italian school of anthropologists (Lombroso and his disciples) are perhaps the most closely akin to those held by Quetelet, but their views no doubt will gradually be changed under the influence of the strong and sober reaction which prevails among the statisticians of the present day.

As remarked above, the calculus of probabilities is one of the main sources of the theory of statistics. This discipline in the past century has had a healthy growth, some of the finest mathematicians having contributed to its development. It is sufficient to name Gauss (method of least squares) and Poisson (*Récherches sur la probabilité des jugements*, 1837). For many years statistics was too little influenced by these contributions. The official statisticians as a rule did not understand or appreciate the law of error. They were content to deal with statistical data in quite an elementary way with the result that much labor was wasted. It was often thought necessary to accumulate immense masses of data in order to prove results which might have been obtained from much less material, if the simple formulas of the calculus of probabilities had been used.

But here, too, came a reaction. One of the leaders was Woolhouse, who published his remarkable article, *On the Philosophy of Statistics* in 1872. Later came the brilliant contributions of F. Y. Edgeworth and of Karl Pearson's school, which took up problems of great significance concerning the physical and moral development of men, particularly with regard to heredity. Important contributions to the theory of statistics were made by the well-known German economist, Lexis, and his disciple, Bortkiewicz. Lexis took up the prob-

lem of regularity in statistical data, dealing principally with the ratio of males to females among new-born children.

A prominent feature of modern statistics is the immense development of the field. The progress of labor statistics during the last few decades is an instance of such extension of official statistical work. A further broadening of the field is due to the private efforts of pioneers, supported by the generous gifts of men like Francis Galton, and to the activities of life insurance and accident insurance companies and institutions investigating the mortality of the insured, or the frequency of invalidity, illness, or accident.

In this extension of the field of statistics much has depended on the attitude of the people towards statistical inquiries. The spread of education through good public schools has been an important factor in allaying popular suspicion. The former distrust lest statistical inquiries might prove to be a preliminary to increased taxes has changed into a general confidence and a willingness to coöperate with the authorities. Consequently in the future there will be rather an *embarras de richesses;* the difficulty will be not so much in gathering material as in mastering it, in digesting all these masses of reports which have been stored in the archives and on the book-shelves of statistical offices.

II.

What is the present outlook for statistics? Some hints of future possibilities are suggested by the foregoing outline of its development.

Let us consider first the future extension of the field of statistics. I have already mentioned the immense progress which has been made of late. It might perhaps seem difficult to make further advances, but the movement has been so rapid, that we can hardly think it will at once come to a standstill, and at all events some opportunities of further progress are evident. We need only to follow the lead of the old political arithmeticians. An essential feature in those early days was the frequent use of representative statistics, where observations covering the whole ground were not available. This method was frequently used too naïvely, but its principle

was not bad. Laplace pointed in the right direction, claiming that the representative districts should be very numerous and spread over the whole area. What remains to be done, is simply to develop a theory of these representative enumerations, stating the limits of the deviations from the true proportions and showing how to approach as nearly as possible to the truth. In fact in many cases it will be practically impossible to do without representative statistics. A typical instance is the production of milk in a country. No statistician would think of measuring the whole quantity of milk, yielded by all the cows in the country. It is only necessary to select a certain number of cows, fairly representative of the whole, and to test the quantity of milk from these cows on certain days. From these observations the whole product may be approximately estimated. So also with regard to the production of butter, meat, etc. The statistics of crops, too, will always depend, somehow or other, on representative data. Again, we may ask what is the amount of timber, which can be produced in a forest: we can only solve the problem by choosing certain select areas in the forest and proceeding to measure the timber in each. The rates to be charged by insurance offices is another example. By taking samples of workshops, cotton mills, theaters or farm buildings, we can find their characteristic fire risk, and from these observations we can derive the rates for similar buildings. Interesting examples of representative statistics are furnished by the Norwegian investigations into the distribution of incomes. By selecting certain places, certain streets and houses, and inquiring about the income and other circumstances of each person there, more detailed information could be secured than from returns covering a whole community, though in fact the modern systems of income taxes have made such knowledge easier to get than it used to be.

Related to the method of representative statistics are certain methods of interpolation. Supposing that the distribution of a certain population according to years of age is known, but that in dividing the population into groups according to occupation it was found too expensive, in spite of the enormous technical progress of recent decades,—

electric counting machines, etc.,—to give the same details, so that the occupation classes are divided only into broad age-groups, then we can fairly well calculate the detailed age distribution, on the basis of the age distribution in the whole population. The formula of interpolation applied may be a rather crude one, as was the case with Süssmilch's estimate of infant mortality in one class of population on the basis of that of another, or we may invent more refined methods; at all events the principle will be the same. As an instance we may take the comparative mortality among widowers and and married men. If the numbers living are classified in ten-year age groups, 55–64, 65–74, etc., it is clear that in each group the mean age of widowers will be higher than the mean age of married men. For this reason there will be proportionally more deaths of widowers, even if their mortality is just the same, and we will be tempted to draw erroneous conclusions from the observations. By means of interpolation we can reduce the age periods and thus remove the described source of error.

III.

Another group of most important problems relates to the correctness of the numbers observed.

Statistical observations are undoubtedly much more correct nowadays than formerly. But of course these observations are not and cannot be entirely accurate. In enumerating a population at a given moment we can get fairly close to the truth, but it will be hardly possible to register all the inhabitants without any omissions (for instance tramps) or altogether to avoid double counting. There will always be some little possibility of error left, and the question is, whether the limits of such errors and the effect they may have on our conclusions can be determined. Here we can learn from the old political arithmetic, for these authors, naïve and unpractical as they often were, frequently had a clear understanding of the importance of finding the limits of error in the statistical material. Süssmilch gives a good instance with regard to Berlin, Heysham with regard to Carlisle, and Milne's correspondence with Heysham concerning the Carlisle observations testifies to his anxious endeavor to

correct his materials as much as possible before calculating his mortality table. This ought to be a characteristic of every statistical report; the degree of accuracy of the observations should always be carefully considered. It would be a useful task to gather observations on the limits of error out of the enormous literature of statistical reports, and to try to systematize the results. Such a work would probably contribute not a little to our confidence in numerical data and would help us to draw valid conclusions in spite of the defects of the materials. There are in fact already available not a few discussions of this kind and undoubtedly more will appear as soon as the question has attracted the attention of official statisticians.

In several cases we have ways of testing statistical data. Sometimes certain questions are answered from two different points of view: for instance, in labor statistics, when reports are secured both from employers and from trade unions. In statistics of international trade the import of goods from one country into another should correspond to the export of the same goods from the former country into the latter.

Frequently it will not be possible without extra effort to determine the margin of error. For example, in age returns at some periods of life there is a temptation to overstate age: an infant 11 months old may perhaps be registered as having completed the first year of life; in extreme old age many persons, proud of their age, are tempted to exaggerate it a little. Again the gentler sex is often suspected of understating age. Further, there is the concentration on multiples of five or ten, common in census reports even where not only the age in years but also the date of birth is entered in the schedules; for many persons, ignorant of their exact age, are inclined to state it in years as a multiple of five or ten and then compute the year of birth by subtraction from the census year. A revision of the census figures is therefore necessary. In some cases a complete revision is possible. Thus in Norway, the country of centenarians, a careful revision of the data concerning persons in extreme old age was undertaken by comparison with the parish registers of births. In other cases we must confine ourselves to making a thorough investigation

of representative or select parts of the material, for instance certain parishes; from these parts it is often possible to draw valid conclusions applicable to the whole.

A certain pessimism is often encountered among official statisticians or economists who attempt to draw conclusions from statistics. It is maintained that the inaccuracies are often so great that it is impossible to get any reliable results. Here, then, is another important problem in the theory of statistics, viz., to determine the significance of the inaccuracies, to state to what extent it is possible to draw conclusions from statistical data, in spite of their imperfections. Here it would be most useful to form a theory of the applicability of imperfect data. For it can easily be shown that even extremely incorrect data may under certain circumstances allow of perfectly safe conclusions. It will be sufficient to discuss a few examples.

Let us suppose that we want to find the mortality of hospital patients suffering from pneumonia, grouped according to our estimate of their temperance or intemperance in the use of alcoholic beverages. Here the division line is very vague. Two persons with just the same habits may be registered, one as temperate, one as intemperate. The pessimists would maintain that observations of this kind are wholly without statistical value. But there is no reason to think that the observers will knowingly falsify the facts. There are patients whom every observer will judge intemperate and others who are as clearly temperate; between these two groups there are some whom it may be difficult to assign definitely to either group, and here different observers might make different decisions. Now let us further suppose that persons with intemperate habits suffering from pneumonia have a death rate of 40 per cent., whereas patients whose consumption of spirits is under a certain limit have a mortality of only 30 per cent., then as a consequence of the error in classification, the results may not show the true difference in relative mortality of the two groups, but it can easily be shown that the results will always indicate in which group the mortality is higher. If out of 1,000 patients recorded as drunkards only 900 are really such, whereas 100 ought to have been registered as temperate, and

a similar group of 1,000 "temperate" patients contains 100 who are really intemperate, then the observed number of deaths in the former group would be, not 400, but only 360+30=390, whereas in the latter group mortality would be 270+40=310 instead of 300. The difference, still marked, is merely somewhat reduced. If mistakes were still more frequent, so that, for instance, 400 of one group ought to have been registered in the other one and *vice versa*, then mortality would have been 240+120=360 and 180+160=340; there would still be a difference left. Only if more were reported incorrectly than correctly, would we get an erroneous result; where this is not the case, we find a difference testifying to an unfavorable effect of intemperance, even if we cannot measure this effect; we can only say that the difference is at all events not smaller than that which we have observed. But in statistics it often matters little whether we can measure the true effect of the causes which are acting or not, if we can only prove that such causes are in existence.

Or how does the health of children in parish-schools vary according to the economic circumstances of the parents? It may be that one observer will consider as belonging to a poor home a child whom another would report as well-to-do. Thus we have some chance of error in the observations, a fact, however, which does not prevent us from finding a difference in health among the two groups, if poverty has a deteriorating effect on health.

Or we may take a typical text-book example. It is a well-known fact that the death rates of legitimate and illegitimate infants at different ages apparently follow different laws. Up to a certain age, legitimate children have a lower rate of mortality than illegitimate, but after that age, a higher rate. In former days statisticians were often inclined to ascribe this phenomenon to a selection among illegitimate children, the more delicate being overtaken by an early death, whereas the healthier children survived. Thus the stock gradually grew healthier, whereas among children born in wedlock the opposite would be the case, delicate children being kept artificially alive, till at last they succumbed in spite of all the care of their parents. But there is an evident source of error,

viz., the legitimation of children by the subsequent marriage, of their parents. If such a child dies after the marriage, it will generally be registered as legitimate, though at birth it was reported as illegitimate. Consequently the fraction measuring the mortality of the illegitimate children is too small, of the other group, too large. As long as the mortality among legitimate children is smaller than among illegitimate, we know that there is an unfavorable cause acting on the latter and that the effect is even greater than the observed difference; but when the difference is negative, we are unable without further data to say whether the difference should be ascribed to defects in the data or to a selective process. We owe to the German statistician, Boeckh, a series of excellent investigations on this subject, proving that legitimate children in Berlin at no age of life had a higher mortality then illegitimate children.

This example is another illustration of the fact that we are often unable to measure the exact numerical effect of a cause, though we can prove that it really exerts some influence. But often it is unnecessary to know more than that. If we have found that a certain cause is detrimental to health, we may leave it to social legislation or to hygiene to bring about a change if possible, to secure which it is not so important to know the exact statistical quantities.

IV.

A third problem of wide import relates to the "law of error." We have seen that the old political arithmeticians did not realize the necessity of investigating the limits within which the observed values deviate from the standard, and, notwithstanding several efforts in this field of late years, much work has yet to be done. Outside of anthropometry, which presents interesting analogies to the law of error, not much had been done until Lexis took up the question, as mentioned above; Bortkiewicz made interesting investigations, particularly with regard to small numbers of observations (*Das Gesetz der kleinen Zahlen*, 1898); and I may refer to my own investigations beginning as far back as 1884.

Thus, in trying to explain the differences in the mortality, of married men and widowers, we might ask whether by chance there had been an uncommonly large number of widowers with ill-health, or living under such circumstances that they were more liable to attacks from epidemic diseases, or perhaps with a very small income, etc.,—circumstances, which have no connection with conditions of married or unmarried life. If this were the case, another investigation might give an opposite result and show a greater mortality among married men than among widowers.

In order to find the limits of these chance deviations, we may first consult the theory of the calculus of probabilities, asking whether Bernoulli's theorem will give us a point of departure. Where the theorems of the calculus of probabilities hold good, the observations will be grouped around the standard value according to the binomial law, found by expanding $(p+q)^n$, where p stands for the frequency of the event, q for its complement $(p+q=1)$ and where n is the total number of cases. Where n is sufficiently large, we can approximately measure the deviations from the standard by the "mean error" $(\sqrt{npq})$. In two cases out of three the deviation from the standard will be smaller than the mean error and in 19 cases out of 20 smaller than twice this quantity, and a deviation over four times the mean error will be exceedingly rare. Simple as this formula is, we can simplify it still more, if the value of p or q is not very great. If $p=0.1$ and $n=10{,}000$, we have a mean error, $\sqrt{10{,}000\times0.1\times0.9}=30$. If we disregard $q=0.9$ in the formula, the mean error will be only a little more; namely, $\sqrt{1{,}000}=31.6$; we have then only to take the square root of the number of deaths (np) as a standard of the deviations. Practically this change makes little difference, and where the value of p is still smaller, the difference will often be very small. Thus, if $p=0.01$, $n=1{,}000{,}000$, we have for the mean error $\sqrt{990}$ or 31.5, but leaving out $q=0.99$, we get $\sqrt{1{,}000}=31.6$, or a quite insignificant difference.

The use of this formula is evident. If the observed number of deaths is 1,020 or 1,030 instead of 1,000, we may look upon it as an ordinary deviation, which need not be ascribed to peculiar causes. But should the deaths increase for instance

to 1,100 or 1,200, we should naturally infer that some peculiar causes have been acting, and the next step will be to make further observations in order to get a clearer understanding of the matter.

The same formula in a little more complicated form can be applied to the chief problem in medical statistics; viz., to find whether a particular method of treatment of a disease is effective. Let the mortality of patients suffering from the disease be p_2, when treated with a serum, p_1, when treated without it, and let the numbers of observations in each case be n_2 and n_1. Then the mean error of the difference between the frequencies of dying in the two groups will be $\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$, and we can get an approximation by putting the observed relative values instead of p_1 and p_2. If, for instance, 200 patients have been treated in each group, the numbers of deaths being 40 and 80, the observed difference thus being 0.20, we shall have the mean error of the difference $\sqrt{0.008+0.0012}=0.045$. The observed difference is four to five times the mean error, and we are consequently justified in believing in the favorable effect of the treatment by serum.

But the question arises: is this test applicable, or do the casual deviations follow some other law? As mentioned above, the students of the calculus of probabilities were often mistaken, for explicitly or implicitly they considered the theorems as axioms, which would hold good at least in all games of chance. In fact this can only be tested by experience. Although d'Alembert in his controversy with other mathematicians had the worst of it, he was right in wishing observations; he had actually touched a weak spot in the theory.

It may, however, be considered as sufficiently proved by experience that the calculus of probabilities can be applied to the problems of games and similar questions. How shall we explain this result? If we toss a coin in the air, or throw dice on a table, there will always be numerous circumstances which we cannot register at all, but which combined contribute to the definite result of the games. The die is on the table in a certain position and at a certain distance from the player; the position of his hand varies; the die is thrown with varying

force; etc. If all the circumstances each time were exactly alike, then the same thing would always happen; if the die had once shown a six, the same number would be shown at every throw; if one card had once been drawn, the same card would constantly appear, etc. In fact all these innumerable individual causes coöperate, so to speak, so that their effect partially disappears, and we are consequently enabled, within the limits given by the "law of error," to foretell the collective result of a number of experiments. We cannot tell what number the die will show at a single throw; but if we repeat the throw, for instance 12,000 times, we know from experience that about 2,000 of these cases will give sixes, and we can with the help of the "mean error" easily find the limits of the possible deviation from this standard. The mean error is 41, consequently we will very rarely expect a deviation of, say 120 or 160; if the die should give sixes 2,200 times instead of 2,000, we should certainly look upon it as "false"; it must have had some quality favoring this event, possibly a certain part of it was heavier than a corresponding part of other dice. If, then, we have found sixes 2,200 times, we know that there is a peculiar cause which we shall have to find in some way or other. The greater the difference, the easier will it be to find the cause.

Now, in vital or economic statistics we have an analogy to these phenomena. Here, too, a certain regularity prevails; particular individual causes are, so to speak, eliminated in the collective results. No two persons are alike; one will be attacked by an epidemic disease, another will escape; one will marry, another remain single; etc., but on the whole in a certain class a certain number of deaths or marriages will take place. The question now is, What are the limits of deviation from the average? Can we find a conformity to the binomial law of error similar to that found in playing at dice or cards, or do other laws of frequency apply to these phenomena?

In vital or economic statistics most numbers have a much wider margin of deviation than is experienced in games. Thus the death rate, the birth rate, the marriage rate, or the relative frequency of suicide fluctuates within wide limits. But it can be proved that, by dividing the observations, sooner or later a marked tendency to the binomial law is revealed in

some part of the observations. Thus, the birth rate varies greatly from year to year; but every year nearly the same ratio between boys and girls, and the same proportions of stillbirths, and of twins are observed. The causes which bring about an increase or decrease in the number of births thus have apparently the same influence on both sexes.

A curious illustration of this is found in the statistics of divorces in Berlin, 1899–1908.

Year.	Number of Divorces.			Percentage of Divorces.	
	Total.	Protestant.	Mixed Religious Belief.	Protestant.	Mixed.
1899	1,608	1,261	232	78	14
1900	936	732	128	78	14
1901	984	769	149	78	15
1902	1,227	972	174	79	14
1903	1,269	981	198	77	16
1904	1,376	1,048	229	76	17
1905	1,421	1,103	210	78	15
1906	1,639	1,271	266	78	16
1907	1,781	1,350	298	76	17
1908	1,868	1,444	291	77	16
	14,109	10,931	2,184	77	15

It appears from this table that the absolute number of divorces in Berlin has fluctuated widely. A change in the divorce law accounts for the low number in 1900 compared with that of 1899, but except for this the numbers have been constantly on the increase. The distribution according to religious belief, however, is very regular; calculating the mean error of the percentage numbers, we find that the fluctuations are within the limits of the binomial law. Hence we may reasonably conclude that the causes affecting the numbers of divorces act with the same force on the two groups of marriages; the individual causes affecting the percentages in each group have collective results of the same nature as in the case of games.

It is easy to multiply illustrations of this kind, showing the grouping of relative numbers about a standard. By the test of experience, then, it is possible to apply the binomial law of frequency in many fields of vital statistics. It seems more difficult to apply the law to absolute numbers, calculating for instance, the number of births or deaths in a given year,

for the fluctuations of these numbers generally exceed the limits of the binomial law. Still we can sometimes find the causes of variations and thereby calculate the absolute numbers more precisely. As regards the number of births it is necessary to make allowance for a tendency to a decrease of fertility; the number of legitimate births depends on the number of married couples, the age of the parties, and the duration of their marriage. A calculation for Denmark through a period when there was no noticeable tendency to a decline of fertility, taking into consideration only the fertility of married women by five-year age groups, in urban and rural districts, but without making allowances for duration of marriage or age of husbands, gave a favorable result, showing that it was not impossible to calculate with tolerable accuracy the number of legitimate births in a particular year from the experience of previous years.

As to the deaths it seems more difficult to calculate the absolute numbers, because meteorological and economic causes exert an influence the numerical effect of which it is hardly possible to measure. By leaving out of consideration ages at which variations in climatic conditions cause great fluctuations in deaths (infancy), we may make some advance towards the binomial law. But there is no difficulty in getting several important results concerning relative numbers. The level of mortality may be very different from year to year, but we can perceive a tendency to the binomial law in the relative numbers, the death rates by age, sex, occupation, etc. By observing the death rate of two occupations in the same period and in the same country we may find that mortality in these professions is rising or sinking with the general level, and we are in a position to conclude from the figures which occupation is more healthful.

So, too, in the case of divorces. The absolute number varies greatly, as we have seen, but the distribution according to religious belief is rather regular. Knowing that 3 per cent. of the dissolved marriages in Berlin are Catholic, whereas about 5 per cent. of the weddings are Catholic, we are warranted in inferring that Catholic marriages are less likely to be terminated by divorce than others.

In economic statistics it is more difficult to apply the binomial law, the conditions being as a rule much more complicated. But if we carefully single out homogeneous observations, we shall often find a decided tendency towards the binomial law. I may refer to interesting examples in H. L. Moore's work, *Laws of Wages* (1911). Another instance can be taken from the Copenhagen statistics of wages (1909). We have here the following distribution per 1,000:

Range of Yearly Income in Crowns.	Wage-earners per 1,000.	Range of Yearly Income in Crowns.	Wage-earners per 1,000.
Under 100	7	Under 900	416
100–200	24	900–1,000	18
200–300	37	1,000–1,200	48
300–400	40	1,200–1,400	139
400–500	52	1,400–1,600	167
500–600	62	1,600–1,800	119
600–700	70	1,800–2,000	77
700–800	66	Above 2,000	16
800–900	58	Total	1,000
Under 900	416		

This distribution has little resemblance to the binomial law: it has two maxima. But by dividing the observations into several classes we may discover a tendency towards the law. Thus, we can ask how each occupation is grouped around its special average, how many persons have, for instance, at least 100 or 200 crowns yearly more than this average, etc. If we then displace all these series of observations so that all the centers coincide, we shall find how all these working men and women together are grouped around their averages, and we may compare this to the exponential law (which closely resembles the binomial law). The result will be as follows:

Deviation from Average	Frequency.		Frequency Expected According to Exponential Law.
	Under the Average.	Above the Average.	
(Crowns).			
0–100	223	212	195
100–200	154	148	152
200–300	84	73	91
300–400	33	29	42
400–500	6	19	15
500–600	3	9	4
600–700	2	1	1
700–800	2	0	0
800–900	2	0	0
Total	509	491	500

On the whole it may be said that there is a tendency toward the exponential law; and this will hold true, if we deal separately with males and females, with skilled and unskilled workers. If some few occupations show conspicuous differences from the law, it always seems possible to find the reason. Thus, in gasworks there seem to be two distinct groups of laborers, each group with its own average.

In anthropometry we have a good analogy to the theorems of the calculus of probabilities. The distribution of height among young men registered for military training is an excellent example. This analogy enables us to find interesting results with regard to typical differences between rural and urban population, rich and poor, etc.

If no conformity to the binomial law can be found in a series of observations, the first task is the classification of the data. The most obvious classifications are by sex, age, profession, residence, etc.

If we do not at once reach the binomial law, we must try further subdivisions. Since each of these classes or subclasses has its peculiar conditions, and is under the influence of different systems of causes, we can say that the more quickly we reach the point where the binomial law holds good, the fewer causes have been acting, whereas the greater the deviation from the binomial law, the greater the number of active causes. An instructive instance of this is found in the discussion of the balance of male and female births—one of the most interesting chapters of the history of statistics. According to Hofacker and Sadler, if proportionally many young men died—as in the present war—then nature would try to restore the balance, so that the proportionate number of male births would increase; couples where the wives were much younger than their husbands would have a relatively numerous male issue. The difficulty here is simply that we generally find very soon a conformity to the binomial law. Moreover, statisticians who dealt with this problem frequently misunderstood the results; in fact very often they had not observations enough, the differences they found being within the limits of the law of error. The late Dr. Geissler found two sets of causes acting against each other and thus creating a tendency to balance:

there are families with relatively many chances of getting children of one sex only, but where this uniformity is once broken, then there is greater probability in future of getting children of the other sex.

V.

As far as I can see, the typical frequency curve in all vital, social, or economic statistics is always the binomial one; but it will require much investigation by statisticians to get to the bottom of this question, to prove whether this supposition is right, or under what conditions the observations will show a tendency to the binomial law. This third problem of the theory of statistics will require much patient labor in careful analysis of each separate class of material. If the observations show the validity of the binomial law, then we know that we can use the above described test of the mean error with little or no hesitation and that on the whole the calculus of probabilities is applicable.

Many statisticians have tried to apply other frequency curves than those with which I have here dealt. We are indebted especially to Karl Pearson and his disciples for important investigations on this question, leading, for instance, to curves with "skew variations," etc. The binomial law is generally a little askew, but for all practical purposes we may set aside this small difference from the symmetric form. Very often, however, we meet with distributions of quite another form; how are we to deal with such phenomena?

I may here insert that it is not as a rule necessary to know the mathematical formulas of these distributions. We may deal with them well enough, if we only know by experience the form of the curves. Distributions according to the binomial or the exponential law need no strict mathematical treatment, the observations from games will give us sufficient materials for finding the form of the curves rather accurately, even for more complicated data. A curious instance is presented by the Danish statistics of fertility. If marriages of 10–15 years' duration are classified according to social position, certain classes, represented by 3,759 families, are found where the birth rate is high, and others, represented by 2,082 families, where it has declined to a low figure. If the couples in both

groups are distributed according to size of family, after a certain size of family is reached, we find that the numbers in both groups show a close correspondence. Out of 2,082 families with low fertility, 192 had more than six children each, distributed in practically the same way as the corresponding 753 couples in the group with a high birth rate. But these 753 correspond to 3,006 couples in the group with the high birth rate with six children or less; in the group with a low birth rate we may separate out a corresponding number, 766 couples with six children or less,—this number bearing the same proportion to 192 as 3,006 to 753. Assuming that these 766 are distributed as in the larger group, we have the following figures:

Number of Children	958 Couples with a High Birth Rate	1,124 Couples with a Low Birth Rate	The Whole Group
0	94	214	308
1	60	180	240
2	99	260	359
3	108	187	295
4	128	196	324
5	149	55	204
6	128	32	160
Under six	766	1,124	1,890
7	102	..	102
8	51	..	51
9	25	..	25
10 and over	14	..	14
More than six	192	..	192
Total	958	1,124	2,082

We have thus found that 958 couples seem not to have limited the number of their offspring, whereas the remaining 1,124 couples appear to have had modern ideas. This is of course only a preliminary investigation. In reality we should have split the group into several classes, each with its own birth statistics. At all events it appears from these numbers, that the movement to limit the birth rate had reached more than half of the group concerned.

It is apparent in a case like this that it is unnecessary to find a mathematical expression for the distribution of families in the two groups. Nor will it be advisable. On account of the great rapidity of the decline of the birth rate the distribution

of the groups will undergo considerable changes, which cannot be foreseen in advance, at least as regards the group with a low birth rate. The curves which we should find would therefore not be typical, and it would be of no use to make the calculations.

But the same objection can be made to several other curves designed to express the distribution of observations in some field or other. One of the most interesting examples is that of the distribution of income among the inhabitants of a country. In trying to draw such a curve we shall certainly find a very skew distribution, the frequency curve differing immensely from the curve conforming to the binomial law. But if we try to divide society into different classes, we shall find in each one a separate distribution with its own center of gravity. In Danish income statistics the figures show a marked tendency for the separate groups to be distributed about an average in a way resembling the binomial distribution; but, if all the groups are taken together, the curve for the whole population assumes a "pyramidal" income distribution (Pareto's law). It is not safe to stop here; a simple change in the constitution of the society, a labor conflict resulting in higher wages, for instance, may give the "pyramid" another form. This will be the case, too, if the commercial classes get higher profits from their business, or if the share of rent decreases or increases, etc.; all these causes are to a certain degree connected with one another, but so much depends on unknown factors that we can not well foresee the changes of the figures. Taking each class alone, we have a much better foundation for our investigation, and then we can return to the total population to get a view of the collective results of all these peculiar movements. Using class statistics as a starting point, we are in fact much nearer the significant causes. If we seek a formula for the combined effects of all the causes in action, we run the risk of overlooking some, which it would really be exceedingly important to take into consideration. Much labor can indeed be lost in this way.

The number of sick days after an accident affords another instance. As there are always proportionally very few cases of protracted illness after an accident, the result will again be

a very skew variation or a pyramidal distribution. But instead of trying to give this distribution a mathematical expression it would pay better to classify accidents in special groups, each with it peculiar center of gravity.

The life table is another example. Here we have a combined expression necessary for practical use in life insurance, for calculating the present value of future assets and liabilities. But several mathematicians have erred in thinking that it would be possible to find a mathematical law of mortality, a physiological law, as it were. We have several formulas of this kind, by Lambert, Moser, Gompertz, Makeham. For a a certain period of life Makeham's formula is exceedingly practical, but after all it is only a beautiful formula of interpolation. It does not hold good at the end of life, for after the age of 80 mortality seems to vary according to another law, nor does it hold good in infancy. And from 20 to 80 the variation is by no means always as uniform as the formula represents; a new treatment of tuberculosis may cause a diminution of the rates of mortality in a certain period of life, the entrance into married life improves the health, military service causes an increase of mortality, etc. Under certain circumstances there will be even stronger influences, thus, for instance, the mortality of sailors, of persons with inherited diseases. A general mortality table will always be an aggregate of numerous special tables, and, if the groups which form the whole community. undergo change, then the life table itself will change, and it may prove impossible hereafter to apply the formula. For a close study of the causes at work it will be necessary first of all to take into consideration the mortality at each age and in each group.

VI.

Whether or not we use mathematical formulas for these combined expressions there are certain mathematical problems in dealing with these complex averages which a statistician must be able to master.

In trying to get comparable data it is often necessary to go through much preparatory labor. If the numerator and denominator in the fraction measuring the frequency of an event,

are not quite homogeneous, it may be necessary to make certain interpolations. We may have to find by interpolation from the census figures taken every tenth year the number of persons at a given age, who are exposed to the risk of death at a certain moment. On the whole we may here lay down the principle that this interpolation first of all should keep as near real life as possible, taking everything into consideration which might prove of importance, such as the fluctuations in emigration or immigration (as far as we can ascertain them), or of births and deaths. Complicated as these problems frequently are, it may be maintained that on the whole they have been solved with sufficiently close approximation. In fact, in many cases the different methods of interpolation lead to nearly identical results. This is true also with regard to adjustment or graduation, a process closely related to interpolation. An experienced mathematician as a rule will get nearly the same results whether he adjusts the figures graphically, "mechanically," or by some mathematical formula, with or without application of the method of least squares. The possible progress in this field will be therefore chiefly that of refinement, smoothing the values to make the changes appear as regular and even as possible. But on the whole, the theory of statistics cannot expect much improvement from such efforts beyond what has already been achieved.

The best means of treating the changes of a variable in a series of observations is the method used by Daniel Bernoulli in the eighteenth century in studying the influence of smallpox on mortality and further developed by Duvillard at the beginning of the nineteenth century. They used the fiction that all variations were continuous, so that it was possible to use the differential calculus instead of operating with finite differences. This method is of great value, as it simplifies many problems extremely; on the whole, it seems to have been generally adopted, though not always without hesitation.

According to this method we calculate, for instance, the force of mortality, a quantity measuring the risk of dying in the next infinitely short moment. This quantity being μ_x, where x is the age, we have as the number of deaths during a moment of length d_x, with l_x standing for the number exposed

to risk, $\mu_x l_x d_x$. If l_x signifies the number surviving at age x according to a life table, then the number surviving at age l_x+d_x will be $l_x-\frac{dl_x d_x}{d_x}$. The number of deaths during the moment d_x being also $\mu_x l_x d_x$, we have $\mu_x l_x = -\frac{dl_x}{d_x}$. If we are warranted in making certain simple suppositions with regard to the force of mortality, this equation will lead to simple expressions for l_x. But not only can we use the force of mortality as it stands, we can also split it into its elements, finding, for instance, the rate of mortality from a single cause of death, or a group of causes, and further asking what will be the result if one or more causes disappear. Or we may introduce another quantity, the rate of marriage, corresponding to the force of mortality, and ask how many persons will survive as single, how many will marry within a certain interval, how many married people will die in the same period, etc. It will cause no difficulty, if the force of mortality is supposed to be different among bachelors and married persons. The same methods can be applied to other problems, for instance, finding the number of persons who become invalids or who recover or who die after recovery. Here, too, we can use different values for the rate of mortality and allow for the influence of age and the duration of invalidity. Again, we can determine the probability of a person being sentenced for a crime, once, twice, etc. The problems are everywhere of the same kind and they present no particular theoretical difficulty. In fact, the main difficulty at present seems to consist in providing sufficiently correct and complete observations rather than in dealing with the materials afterwards.

VII.

Where are we at present with all these problems? We might describe the program of vital statistics as the task of following human beings from the cradle to the grave, taking account of all their chances, their good fortune and mishaps, their physical and intellectual growth, their criminality and moral defects, etc., etc. What will be a young man's choice

of profession? Will he follow in the steps of his father or not? What will be his chance of getting a better position in society or the risk of his being forced downwards in the struggle? What will be the danger of unemployment, and, after having lost a position, what is the chance of finding work again? What income will a man on an average earn in each occupation and at different ages, and what will be his economic position on reaching old age? Again, will education pay in the competition of life, will a boy who begins as apprentice in some trade get a substantial pecuniary advantage, or will he be actually no better off than an unskilled laborer?

It will be readily seen that there is still a long way to go. Thus, as to birth statistics, we shall have to add considerably to the present store of statistical data on fertility in different classes of society as influenced by the duration of marriage and the age of consorts. In marriage statistics we shall have to observe the different rate in different classes with allowance for the fact that very often a wedding is registered for two places of residence—that of the bride and of the bridegroom—and that the wedding frequently is followed by a change of occupation.

As to mortality statistics there used to be an enormous confusion. In the middle of the past century medical statistics suffered severely from dilettanteism, especially as regards the correlation between mortality and occupation. Much was accomplished in the way of clearing up the subject in England by the efforts of William Farr and his successors, but there is still a serious lack of material. The same may be said of statistics of sickness and invalidity, though observations during a century of experience in Friendly Societies are available and German compulsory insurance against invalidity, with its complicated machinery, has been in force for 25 years. Other problems have hardly been touched as yet, thus the mortality of persons suffering from certain diseases, such as syphilis, mental diseases, etc., the mortality of persons rejected by life insurance companies, the influence of heredity, etc.

Still less cultivated is the field of migration statistics. We have some data with regard to transatlantic migration, but not much more. Still it would not be impossible to find approxi-

mately the rate of migration for different ages (corresponding to rate of mortality). If we know from two consecutive censuses the numbers of persons born in a certain year, and if we know further the rate of mortality, we can find approximately the rate of migration (excess of emigration above immigration); and knowing in some places the numbers of those who immigrated a short time before the census, we have some data by which we can separate this rate of migration into its elements.

Still more complicated are the problems of economic statistics, for here we have frequently one more dimension. We have to deal not only with numbers or quantities, but also with their value in money. Only a very small part of the work has been done, and the scanty materials we have are frequently of doubtful value, because it has rarely been possible to separate them in such a way that the numbers are quite homogeneous. This is the main reason why statistics have so often been considered unreliable.

In agricultural and industrial statistics we have above all the problem of finding the quantity produced and its value in money; as mentioned above, representative investigations in this field are particularly useful or even indispensable. Other important problems to be investigated are the distribution of enterprises by size of farm or industrial establishment, and the working of the law of increasing returns in industry. Statistics of international commerce are somewhat developed, though there are evident defects in the returns, but of internal commerce we know practically nothing. The variations of price under the influence of supply and demand form another series of important statistical problems. Again what is the rapidity of circulation of a coin or a bank-note as it passes from hand to hand, and how can we get observations on the influence of bank reserves or of the rapidity of circulation on prices and the rate of interest? As a link between vital and economic statistics we have, moreover, the question as to the yearly income of a whole nation or a class of population, the statistics of public finances, shifting and incidence of taxation, etc. Even here only very few of the important problems have been thoroughly dealt with.

VIII.

It appears from this sketch of future problems, the list of which any economist might easily extend, that statisticians of the present day have no lack of work. But frequently we observe among statisticians, a certain despair, of their ability not in the collection but in the interpretation of statistical data. In consequence of this pessimism there are statisticians who confine themselves to giving a true description of the phenomena. But we must not forget that a true description cannot be given if we cannot find the relations of cause and effect that prevail, for else we run the risk of giving a confused mass of details instead of a clear outline of the main features of the observations concerned.

It must not be forgotten, however, that the task of the statistician is not so much to find the causality himself as to help others to find it. The statistician must be content if he can show that certain groups of numbers show marked differences, leaving it to physiology, meteorology, and other sciences to explain these differences. Thus his task is indeed a very modest one, but it is always satisfying to be able to put others on the right track or to spare them from some misapprehension of the facts.

Often enough the statisticians will be obliged to acknowledge that there are two or more causes which cannot be separated; it may even be that two causes have a much greater effect if combined than if working separately. So with regard to labor and capital. If a man works without any capital, his produce is quite insignificant; and, if capital is left alone, its product will be zero; no one can tell therefore which part of the produce is to be ascribed to labor and which to capital. A simple example will show how easily confusion arises. Suppose that certain goods to the value of $10,000,000 are exported. If now the exported quantity increases 50 per cent., and prices 20 per cent., then the value of the export will be increased to $18,000,000. If prices were unaltered, but the quantity only changed, then there would be an increase of value of $5,000,000; if prices only varied, we should have $2,000,000 more; thus we could account altogether for $7,000,-

000; the real addition was $8,000,000, or $1,000,000 more. In this case it is easy enough to see what the matter is, but very often the problems are much more complicated. An example from anthropometry will illustrate this.

Let us ask whether height is an inherited characteristic. To make the case as simple as possible, let us suppose that there are in a country two different types, each with its peculiar average, round which the individuals are grouped. A person belonging to the tall type may by chance be of small stature, and *vice versa;* and, if we divide the population into two groups according to height, we shall have a smaller or larger number of the short type in the group of tall men and *vice versa*. Supposing now that the type is inherited, whereas the casual causes of deviations from the typical height are not transmitted to the offspring, we have in the second generation two groups, each numbering persons of both types, but probably with an unequal distribution. Each type having its own center of gravity, the observed average in each group will lie between them, that is to say, the second generation will apparently show a smaller distance between the two sizes than was the case with the fathers. Galton and his disciples would speak here of regression, human beings having apparently a tendency to revert to former types, but it follows from the above that this regression is only a formal effect of the nature of the observations; it does not really lead us to an understanding of the causality.

IX.

Let us now suppose that we have a group of observations in which only one rate of frequency may be said to prevail; then, as shown above, we have the means of controlling our conclusions by finding the limits of deviation from the standard value. This is the most simple of the statistical problems. But generally the data are not homogeneous. Most frequently the material must be split up into several groups, each with its own frequency, and we have then the problem of getting an idea of all these relations together.

We can sometimes confine ourselves to studying each of the groups separately in comparison to other corresponding groups of observations. The two series may stand so clear of each

other that no doubt is possible. When in a certain occupation the mortality in each age group is decidedly greater than in the general population, it is clear that the occupation concerned is under the influence of some unfavorable cause. But frequently the differences are not all in the same direction, sometimes some of the rates in the occupation are lower than in the general population, and the evidence thus is apparently contradictory. This irregularity may perhaps be due to accidental causes, the influence of which would have been eliminated, had the number of observations been larger. It will therefore be necessary to combine these results in order to get a clear idea of the influence prevailing in the occupation. Here we have many methods of procedure. In determining which method is the best, it is a good principle always to keep as near as possible to the original observations, never leaving them out of sight. Taking, in the present case, the mortality of the general population as a standard, we may calculate some expression or other from the two series of observations; the mean duration of life, the value of a life annuity, etc. But, according to my experience, it is always best to choose a method which permits us to keep in view the original numbers. In this way it will be easier to avoid misapprehension and erroneous conclusions.

The simplest standard calculation in the present case is to find the budget of deaths in the occupation,—the deaths that would have occurred in the occupation if the mortality had been that of the general population or other group,—which for some reason we might consider typical. To illustrate this, I shall begin with a very simple case, that of barristers and solicitors according to English vital statistics (1900–1902). The following table gives all the necessary details:

Age Period.	General Population.	Barristers and Solicitors.			
	Death Rate per 1,000.	Population.	Actual Deaths.	Expected Deaths.	Death Rate per 1,000.
25–34	6.38	16,323	77	104	4.72
35–44	10.94	19,686	142	215	7.21
45–54	18.67	13,869	189	259	13.63
55–64	34.80	6,741	188	235	27.89
25–64	14.08	56,619	596	813	10.53

The simplest procedure would be to compare the rate of mortality for barristers and solicitors at all ages (10.53) with that of the general population (14.08), but as the age distribution is not quite the same, this comparison will not be without objection. The rates for each age, however, show very much the same proportion to the corresponding general rates, and we might stop with that; but, if we want a common expression, the budget of deaths will give a practical means of comparison. Under the supposition that the mortality is the same as in the general population, we find in the youngest group 104 expected against 77 actual deaths, in the next age group we have 215 against 142, etc., altogether 813 against 596, the actual mortality being thus only 73 per cent of the expected one. The mean error is 29 or about $^1/_7$ of the deviation (217); the mean error in the single age groups is between $^1/_3$ and $^1/_5$ of the corresponding deviations. The mean error is then proportionally much smaller, if we take the total, than if each group is taken individually. This is a great advantage, for often there are so many age groups (for instance, yearly intervals) that the deviations from the type are very irregular, whereas the total may allow of quite a safe deduction. The separate bits of evidence may be rather uncertain, but if combined they may warrant a definite conclusion.

The calculation of expected deaths will be particularly advisable in case the age distribution of the deaths is unknown. As remarked above, we cannot draw sound conclusions from the crude rate of mortality, but if the expected mortality is calculated, we have in the actual and the expected numbers, two homogeneous quantities, which we can safely compare. In a case like this it may be objected that mortality in one age group very well might be greater than the expected; we cannot tell whether mortality in the age group 25–34 actually has been lower or higher than in the general population. But at any rate it is clear that in some age group or other mortality must have been decidedly lower among barristers than in the general population; we have thus found a valuable result, though it may be worth while to subject it to a closer investigation. But ordinarily one might reasonably assume that the mortality in all age groups is lower, as is actually the case here.

English statisticians often use a modification of the method just described of calculating expected deaths; viz., the method of "standards" (in fact, the method of expected deaths can quite as well claim the name of a "standard" method). The difference between the two forms will be easily understood by using the same example as above. First, let us calculate the number of persons among whom, according to the general table of mortality and with the age distribution of the general population, 100 deaths would occur. Let us further calculate the number of deaths which would take place in this "standard" population, if the mortality were the same as observed among barristers and solicitors. The figures are as follows:

Age Period.	Standard Population.	Mortality per 1,000.		Expected Deaths.	
		General Population.	Barristers and Solicitors.	General Population.	Barristers and Solicitors.
25–34	2,625	6.38	4.72	16.75	12.39
35–44	2,040	10.94	7.21	22.32	14.71
45–54	1,475	18.67	13.63	27.54	20.10
55–64	959	34.80	27.89	33.37	26.75
25–64	7,099	14.08	10.53	100.00	73.95

According to mortality experience among barristers and solicitors, in a standard population approximately 74 would die instead of 100 according to the general rates of mortality. This is very nearly the same proportion as reached above. In the present case the two forms of comparison lead to nearly the same result, and this will generally be the case, if the age distribution in the special group is not much different from that of the general population. But on the whole the method described last is a little more complicated than the calculation of expected deaths, and in particular it is not applicable, if the age distribution of the deaths of the barristers and solicitors is unknown.

Still more complicated are other methods of procedure, such as the calculation of the mean duration of life or of the value of life annuities. In the present case we have the mortality experience for only a part of human life, and for ages above 65 we should be obliged to make estimates. Another way of comparison would be to calculate the number surviving at 65

out of a certain initial population. Such calculations may often throw light on the effect of the peculiar circumstances of the group under investigation; they ought never to be relied upon as the sole method of comparison, but may be used as supplementary to the simpler method of expected deaths.

X.

In economic statistics it will frequently prove of interest to have a common measure, a "standard," in order to get a clear view of the main result of the changes that may have occurred. Thus, when comparing the mercantile marine of two countries, or of the same country at two different periods, it may be useful to assign different weights to steamers and to sailing vessels, multiplying the tonnage of the steamers, for instance, by 3, thus taking into consideration the greater carrying power of steamers. Or in considering the livestock of a country, a cow being counted as 1, we may give a horse the weight of, say, $^3/_2$, a pig as $^1/_4$, a sheep as $^1/_6$, etc., and thus secure a rough estimate of the collective value of the livestock. A calculation of this kind may prove useful, but it cannot replace a close study of the original figures. The calculation may conceal changes which it would be of the highest interest to know thoroughly. Thus, in the history of agriculture, a decrease in the number of sheep might be offset by an increase of swine, and from such a computation we should learn nothing of this change. But having first studied the numbers in detail, we may get a comprehensive view by using some standard calculation. In any case the direct knowledge will always have a greater value than any abstraction, however ingeniously invented; and here, too, it may be laid down as a first principle always to keep the original numbers in sight as far as possible.

A most interesting example of a standard calculation is the index number, the familiar instrument of the economists. In trying to find the purchasing power of money at two different epochs we may calculate the budget of a certain model family, for instance that of a working man with a certain yearly income. If this budget at one time is $500 and the same quantities of all the commodities concerned at a later time would cost, for instance, $550, then the purchasing power

of money for this working man may be said to have decreased in the ratio $^{10}/_{11}$. But, if we used other budgets, we might get other results, so the comparison presents a defect, though probably the difference would generally be small. We are a little further from reality, if we calculate average values of the percentage variations of prices of different commodities, as is the case with most modern calculations of index numbers. In fact, we reduce the problem to that of using a budget in which each commodity enters with unity—or perhaps with a certain weight—a sort of abstraction from the actual budgets used above. It may easily be proved that a calculation of this kind cannot possibly give mathematically exact results. Let prices of two commodities at a given time be equal to unity, let the two prices at a later time be 0.80 and 1.25, and again at a third moment of time unity as before. Now taking our standpoint at the first moment, we find that the general index has increased at the second moment by $^1/_2$ (0.25—0.20) or $2^1/_2$ per cent. But from this moment till the third we have also an increase equal to $\frac{1}{2}$ (0.25—0.20) (the price of one commodity increasing from 0.80 to 1.00, or 25 per cent.). Thus in each of the two intervals prices have increased on an average $2^1/_2$ per cent. But the result of the combined movement from the first moment to the third will be nil, there has in fact been no change whatever. In practice this objection is not of great significance. Most economists would decidedly prefer the simple calculation of general index numbers by taking the arithmetic average of price changes instead of using a geometric average, such as recommended by the late W. Stanley Jevons, whose formula required operations with logarithms instead of with the numbers themselves. But at all events, whether we use one or the other formula, it will be necessary always to keep the various prices themselves in view. In the above example one commodity was becoming cheaper, the other, more expensive, and these facts will require close study before we try to find collective expressions for the changes.

Very often the characters we want to compare are not quantitatively measurable. If we investigate eye color, we may arrange the colors in a scale from blue to brown, and give the first rubric of the scale the weight one, the next two, etc.,

and calculate in this way the "average eye color" of particular groups. Thus, taking an instance from Galton's Family Records, we get the following results with regard to fathers and sons:

	Fathers Grouped According to Eye Color.	Average Number for Sons.
1	36	2.7
2	322	3.1
3	264	3.4
4	180	4.0
5	5	4.2
6	64	4.8
7	101	5.0
8	28	5.6
Total	1,000	3.7

It will be seen from this "contingency" table that the eye color of the sons corresponds to that of the fathers, in that the eye color of the sons shows a constant progression from blue to brown in the same order as that of the fathers, though there is a certain "regression" observable, the dispersion of the sons being much smaller than that of the fathers. Of course there is something unreal in such a comparison, the characters not being quantitatively measurable, and it would be better either to consider only the original data or to modify the calculation so as to conform more clearly to the observations. Professor Pearson proposes a distribution according to the exponential law, assigning to each group a mean distance from the average corresponding to this law. By this method the following distribution is obtained:

Group.	Fathers.	Sons.
1	1.0	3.8
2	3.3	4.3
3	4.8	4.8
4	5.8	5.2
5	6.3	5.6
6	6.5	5.7
7	7.0	5.8
8	8.0	6.4

It will be seen that the numbers on the whole have the same character as before; here, too, there is a smaller disp sion among the sons than might have been expected.

Such a method may be useful in dealing with certain problems, as, for instance, the results of examinations with peculiar marks for each discipline. At all events a detailed study of the reports themselves will always be indispensable as a check upon these calculations; in fact, low marks in one study, combined with very high marks in another, will give a different impression from that of a uniform grade in all studies.

XI.

One of the methods of comparison which have of late become popular among statisticians with mathematical training is that of correlation based on Bravais's formula. A simple example is offered by the age distribution of brides and grooms. If on a piece of plotting paper we put a point for each couple, the age of the bride being the abscissa, that of the bridegroom the ordinate, then we shall find that the points are not spread uniformly over the paper, but that they are clustered near a certain line, forming a sort of milky way. We may then by the method of least squares find the equation of the straight line which may be said to give the best idea of this milky way. This equation being $y-kx=0$, we have $\Sigma(y-kx)^2=$a minimum, which will require $k\Sigma x^2=\Sigma yx$. If we further introduce the expression $\Sigma x^2=ns_x^2$, n being the number of cases, further $\Sigma y^2=ns_y^2$ and $\Sigma xy=ns_{xy}=nrs_{xsy}$, where r is a quantity defined by the equation, we find the equation of the line in question:

$$y=kx=r\frac{s_y}{s_x}x$$

and the sum of the squares of the differences is:

$$\Sigma(y-kx)^2=ns_y^2, \text{ so that we have}$$
$$\mu^2=s_y^2\ (1-r^2).$$

If $r=1$, all the points will lie on the straight line; the smaller r is, the greater will be the mean error; when there is no correlation, we shall have $r=0$. This quantity r is called the coefficient of correlation, whereas the quantity k is called the coefficient of regression.

The formula of correlation will prove useful in all cases where the points are grouped nearly around a straight line, as is the supposition, and this will very often hold good, as in the case supposed of the age distribution of brides and grooms. Still we must not forget that this formula removes us somewhat from the original data and that it does not relieve us from the necessity of making a close investigation of these observations. On the whole, the formula of correlation does not introduce any new principle; by tabulating and grouping the observations we can easily establish as a rule the fact of correlation without the use of the formula. To take an example from Yule's *Introduction to the Theory of Statistics* (1911), the percentage of population in receipt of poor law relief in 38 English Poor Law Unions of an agricultural type is correlated with the average weekly earnings of agricultural laborers, and we find as the coefficient of correlation, $r = -0.66$. But it is unnecessary to make this calculation. Grouping the districts according to percentage of poor law relief, we find the following numbers, which tell us, without any long computation, how relief and wages are related:

Number of Districts.	Per Cent Poor Law Relief.	Average Wages (Shillings).
9	1.90	17.88
10	3.21	16.21
10	4.30	15.00
9	5.27	14.75
38	3.67	15.95

These numbers give us a perfectly clear idea of the connection between wages and poor law relief. The coefficient of correlation will tell us nothing which cannot be seen from an inspection of the original numbers. But in the illustration the amount of poor relief is influenced not solely by the average wages, but also by other influences, for in each group there are conspicuous deviations, and to explain them other causes must be found. The coefficient of correlation teaches us this and no more.

To take another example from Mr. Yule's book, what is the correlation between the fertility of mother and the fertility of

daughters in the British peerage? In each marriage of the mother generation one married daughter was chosen, and the number of her children put down, only marriages with a duration of at least 15 years being taken. It will be seen that the numbers are not quite homogeneous, in the first generation each marriage taken into consideration being fertile and with at least one daughter, in the second generation all marriages being counted.

We have the following numbers:

Mothers.	Daughters.			
Number of Children.	Average Number of Children.	Actual Number.	Expected Number.	Actual Number per 100 Expected Children.
1	3.2	167	230	73
2	3.5	201	247	81
3–4	3.9	905	1,006	90
5–6	4.1	1,087	1,145	95
7–8	4.8	976	889	110
9–10	5.1	655	555	118
Above 10	5.6	344	263	131
Total		4,335	4,335	

It will easily be seen how these calculations have been made; they are in fact quite elementary; the mean error in each group can be found without difficulty. In this method of calculation we have the advantage of being in close contact with the original observations.

Most of the problems which we meet in statistics of families are of the same elementary nature, and we shall gain very little in dealing with them by the more complicated processes. We shall always have to divide the observations in order to find the frequency in question in each group, and, if these groups are too small, we may combine the results by the method of expected cases or by some other simple method of calculation. The dependence of mortality of children on the age of the mother or the number of the birth, etc., or the rate of mortality in "phthisical" families is an example. The main difficulty is perhaps to find the moment in which the persons concerned enter into observation, for instance, when a death from phthisis occurs in the family. The following example is

taken from Lundborg, *Medizinisch-biologische Familienforschungen* (1913), giving details with regard to a certain epileptical disease (myoklonus) found in nine families with altogether 74 children. In this case we can take the family under observation after the birth of the first child suffering from the disease. The method will be seen from the table, in which 20 children who died under the age of ten have been left out of consideration.

Number of Family.	Number of Children in Family.	Number Diseased.	Distribution of Children (the Affected Ones Signified by x).								
			1	2	3	4	5	6	7	8	9
I.	6	3	x	..	..	x	x	..			
II.	8	1	..	x	..	..	..	..	..	..	
III.	6	2	..	x	..	..	..	x			
IV.	9	3	..	..	..	x	x	..	..	x	..
V.	9	1	..	..	..	..	..	..	x	..	..
VI.	5	2	..	x	..	..	x				
VII.	6	2	..	..	x	..	..	x			
VIII.	4	2	x	..	..	x					
IX.	1	1	x								
	54	17									

In the first family the oldest child suffered from the disease; there were five children under observation, two of them being affected; in the next family the second child was diseased, but of the six children born subsequently none was affected, etc. We obtain the following result:

Number of Family.	Number of Children.	Cases of Myoklonus.
I.	5	2
II.	6	0
III.	4	1
IV.	5	2
V.	2	0
VI.	3	1
VII.	3	1
VIII.	3	1
IX.	0	0
	31	8

Thus about one out of four children suffered from this disease. Weinberg proposes another scheme, which involves a somewhat more complicated calculation, leading, however, to substantially the same result.

XII.

Sometimes the problem is complicated by the absence of certain essential data; it is then frequently impossible to draw a safe conclusion. The question of the "handicapping" of first born children having been discussed of late in an interesting article in this Journal* by Louis I. Dublin and Harry Langman, I shall choose that problem as an example. On account of its great importance I shall deal with it at some length, discussing some Danish statistics of persons suffering from tuberculosis.

Number in Family.	Total Children.	Number Affected with Tuberculosis.	Per Cent. Affected.
1	3,522	988	28
2	3,344	713	21
3	3,043	568	18
4	2,619	427	16
5	2,161	271	13
6	1,731	198	11
7	1,296	113	9
8	962	81	8
9	636	46	7
10	445	48	11
11 and over	841	69	8
Total	20,600	3,522	17

Readers of this Journal will easily understand how this table was constructed, by consulting the paper I have cited. Seven hundred and thirteen second born children, for instance, were affected, their first born sisters and brothers will be found among the 3,522 first born, this number also including 988 first born affected with tuberculosis, 568 first born brothers or sisters of third born persons affected, etc.

There is a strong temptation to conclude that the first born, with their 28 per cent. affected have a much greater risk of being attacked by tuberculosis than the second or third, not to mention the later children. But it is evident that these observations are one sided, and that essential details which ought to have been included are omitted from the tabulation. If we inquire what is the probability of children with no sisters or brothers being affected, we find that it is one, all the children

*"On the Handicapping of the First Born" in American Statistical Association, *Quarterly Publications*, XIV (1915), 727-735.

observed in this group being in fact affected. The case resembles very much the problem of constructing a mortality table from observations of deaths only. The present statistics fortunately enough can be separated according to the number of the birth and the size of family, so that we can form the following table:

Number of Affected Child in Family.	Number of Children in Family of the Affected.											
	1	2	3	4	5	6	7	8	9	10	Over 10	Total
1	178	155	170	128	107	74	52	45	26	19	34	988
2	..	146	134	106	100	77	44	45	22	12	27	713
3	..	..	120	109	85	68	56	51	32	19	28	568
4	..	..	..	115	63	81	60	33	31	18	26	427
5	..	..	..	..	75	54	35	38	19	23	27	271
6	..	..	..	..	..	81	43	29	11	12	22	198
7	..	..	..	..	..	..	44	36	13	7	13	113
8	..	..	..	..	..	..	..	49	12	5	15	81
9	..	..	..	..	..	..	..	..	25	6	15	46
10	..	..	..	..	..	..	..	..	..	28	20	48
Over 10	..	..	..	..	..	..	..	..	..	..	69	69
Total	178	301	424	458	430	435	334	326	191	149	296	3,522

It will be seen from this table that 178 affected belonging to families with only one child have to be left entirely out of consideration, there being no other group of observations comparable to them. Consequently we have 810 first born comparable to 713 second born. Likewise we can compare 568 third born to $713-146=567$ second born, etc. We can therefore arrange the observations as follows:

Number in Family.	Number Affected.	Number in Family.	Number Affected.	Expected Numbers.
1	810	2	713	712.1
2	567	3	568	561.6
3	448	4	427	420.3
4	312	5	271	305.8
5	196	6	198	219.8
6	117	7	113	147.3
7	69	8	81	99.6
8	32	9	46	58.8
9	21	10	48	37.6
10 and over	60	11 and over	69	..
Total	2,632		2,534	—

The expected numbers have been calculated by distributing 301 affected persons belonging to families with two children

uniformly on each number, further a third part of 424 persons in families with three children on each of the three birth numbers concerned, and so on, for instance

$$\tfrac{1}{2} \cdot 301 + \tfrac{1}{3} \cdot 424 + \tfrac{1}{4} \cdot 458 + \ldots = 712.1$$

The two columns in the above table would be homogeneous, if there were no other defects (as we shall see directly that there are). On examination the numbers of affected are found to correspond with the expected numbers fairly closely; there is no considerable difference with the single exception of the first born, who indeed seem a little "handicapped" compared to the later children. But the numbers are not quite homogeneous. To show this, let us suppose that the year of observation is 1910 and that each affected person was 30 years old, so that all the persons affected with tuberculosis in the table were born about 1880. Now the marriages of the parents of these persons are from varying epochs. Probably the parents of the first born generally married in 1879, of the sixth born perhaps about 1869. Comparing now in the table on p. 48 the numbers of the first and of the sixth child in families with at least six children, we find 250 against 198. But these 198 belong to marriages which are, say, 10 years older than those to which the 250 first born belong. But since the population has been increasing steadily, we cannot compare the numbers till we have made an allowance for this increase. Thus the difference between the two numbers will be much reduced. It seems probable, however, that we should find some little difference in favor of the later born children; it must be remembered that the first born children generally have a comparatively high infant mortality, so that proportionally few will survive,—a fact the correction for which tends to increase the difference. At all events the "handicapping" of the first born, as far as tuberculosis is concerned, must be insignificant compared to the enormous differences indicated in the table (p. 47).

This example shows how very difficult it is to handle one-sided observations. We are safe only when we have data not only on the numbers affected by tuberculosis but also on the whole population among which these cases of disease have been observed. What is wanted is a fraction, the numerator of

which gives the number of cases of disease, deaths, etc., and the denominator, the total number exposed to risk. Thus only can we draw a safe conclusion. If these required elements are not available, we must find whether and how far the defects in the statistics will invalidate our conclusions, by taking the various defects into consideration, as described above (pp. 17, ff.). This process, as we have seen in the present case, is often a very difficult task.

XIII.

Returning now to the problem of dealing with observations of characters which are not quantitatively measurable, I shall borrow some figures from a paper by Professor Pearson in *Biometrika* (III, 1904). One thousand seventy six pairs of sisters in a school were characterized by the teacher according as they were "quick," "good natured" or "sullen." The figures were as follows:

99 pairs, both sisters quick.
498 pairs, both sisters good natured.
60 pairs, both sisters sullen.
177 pairs, one sister quick, one good natured.
77 pairs, one sister quick, one sullen.
165 pairs, one sister good natured, one sullen.

Here again it seems to me that the data require only an elementary treatment in order to give full evidence as to the influences concerned. We have among 2,152 children 452 reported as quick, 1,338 as good natured and 362 as sullen (or 210, 622, 168 per 1,000): consequently according to the calculus of probabilities 44 per 1,000 of the pairs should be both of them quick, 261 should have one quick and one good natured sister, etc. We can construct the following table:

	Expected Number.	Actual Number.
Both sisters quick	48	99
Both sisters good natured	416	498
Both sisters sullen	30	60
One sister quick, one good natured	281	177
One sister quick, one sullen	76	77
One sister good natured, one sullen	225	165
Total	1,076	1,076

We see at once that the sisters are very often alike, 657 pairs having the same temperament against 494 as expected. It is not difficult to find the mean error in order to judge how much of the difference between the expected and actual cases is due to other causes than chance variation.

Professor Pearson prefers another way of dealing with this problem, arranging the experience in the following table:

Second Sister.	First Sister.			
	Quick.	Good Natured.	Sullen.	Total.
Quick	198	177	77	452
Good natured	177	996	165	1,338
Sullen	77	165	120	362
Total	452	1,338	362	2,152

We find thus each pair twice, in 77 cases one sister is quick, the other sullen; these are tabulated as 77 quick sullen and 77 sullen quick, etc. Giving now each group a separate weight we may calculate a coefficient of correlation. The result will be as above, but the calculations will require much more time than the more elementary procedure, and the results do not present any clearer view of the influences in question. We may simplify the table by reducing the number of columns, for instance, by adding the good natured and sullen, so that we get the following result:

Second Sister.	First Sister.		
	Quick.	Good Natured or Sullen.	Total.
Quick	198	254	452
Good natured or sullen	254	1,446	1,700
Total	452	1,700	2,152

But even thus we shall have more calculations than in the above elementary solution of the problem.

Sometimes even these contingency calculations increase the distance between the original observations and the calculated results to such an extent that we should decidedly

prefer the elementary calculations. Thus, in a recent review in the *Journal of the Royal Statistical Society* (Jan. 1916) of *The Material Culture and Social Institutions of the Simpler Peoples*, by Hobhouse, Wheeler, and Ginsberg, it is objected that a coefficient of contingency ought to have been calculated instead of confining the description of the conditions of these people to the simple statement that 47 per cent. of the "lower Hunters," 25 per cent. of the "higher Hunters," and 10 per cent. of the lower "pastoral" group fall into the category, "government slight or nil." It seems to me that this comparison is quite sufficient and that it would be waste of time to proceed to further calculations.

If I am right in these remarks, statistics of this kind can generally be treated in quite an elementary way, and no new results can be found by using more complicated methods. The principal problem will always be to find the relative frequency of the event or the character which we want to investigate, and first of all our object must be not so much to prepare refined statistical methods as to provide useful observations. The special nature of the observations may then lead to modifications of the formulas, to a more complete and refined theory, but at present we are more in need of statistical data than of theoretical investigations. This may sound curious, in view of the immense mass of details which are published every day by the numerous statistical institutions all over the world, but to a great extent all these reports are repetitions, so to speak, of older investigations, most of them made in a single mold. We want statistical observations covering new fields, and here, as shown above, an enormous amount of work remains to be done. Of course we cannot do without these myriads of statistical volumes; they have, at least the great bulk of them, their local claim to exist;—but beyond these reports numerous problems are waiting to be solved and the solution will require much patience and much careful work in gathering the necessary materials. As far as I can see, the future of statistics will depend on the energy with which these problems are taken up.

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By Warren M. Persons, *Colorado College.*

The reading of Westergaard's article on "Scope and Method of Statistics" makes me hope for the early translation of the manuscript and publication in English of his work on statistical theory. Particularly, I would like to have access to Westergaard's treatment of the probable error in time and geographic series, his discussion of the Pearsonian coefficient of correlation and his use of elementary methods of analysing data.

The present article leaves one doubtful concerning several points. For instance, at the end of Part I Westergaard says, "Consequently, there will be in the future an *embarras de richesses;* the difficulty will be not so much in gathering material as in mastering it and perusing all these masses of reports which have been stored in the archives and on the book-shelves of the statistical offices." In the concluding paragraph of the article he says, "First of all our object must be not so much to prepare refined statistical methods as to provide useful observations . . . at present we are more in need of statistical data than of theoretical investigations." In fact, we need both. If we have at hand the calculus which will enable us to derive the meaning from data of a certain kind, such a calculus will, first, encourage us to secure the necessary data and, second, determine the form of the data to be secured.

Westergaard does not correctly describe the economic statistics available today when he refers to them as "scanty material" of "doubtful reliability." Statistics of interest rates, exchange rates, production of commodities, prices, railroad earnings, clearings and the like are neither scanty nor unreliable. It is true we need more data, well classified; we also need a generally accepted calculus for treating them.

With Westergaard's characterization of the Pearsonian coefficient of correlation I emphatically disagree: "The coefficient of correlation will tell us nothing which cannot be

seen from an inspection of the original numbers." Recently I have been working on a problem involving the following question: Is the maximum correlation between steel production and prices and clearings obtained when concurrent items are paired or for a lag in one of the series? Such a question can be answered only by the application of a device such as the coefficient of correlation which enables us to measure the degree of correlation.

Representative investigations should, as Westergaard says, be used more frequently. Methods of interpolation and the theory of error must of necessity be developed for the treatment of the data thus secured.

My experience accords with Westergaard's that "it is possible to apply the binomial law of frequency to many fields of vital statistics," that in economic statistics the law is more difficult to apply, but that "the typical frequency curve in all vital, social, or economic statistics is always the binomial one." I agree that search for a formula which fits accurately a statistical series is usually unnecessary and that the securing of such a function is of use mainly, if not exclusively, for interpolating values. It is obvious that by using a sufficiently large number of constants a function may be found to "fit" with a high degree of accuracy any statistical series.

Westergaard's caution that we should "keep the original data in view" is one that every investigator should hold in mind.

The publication of Westergaard's book in English will be warmly welcomed by American statisticians.

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By Julius H. Parmelee, *Washington, D. C.*

Westergaard's treatise is an inspiration to statisticians; first, because it breathes a wholesome optimism as to the mission of statistics in the settlement of social problems; secondly, because it offers definite suggestions for the utilization of mathematical laws in statistical work.

The thirteen sections of the treatise have a marked German flavor, in their painstaking thoroughness, their attention to a wide mass of detail, and their fidelity to what may be termed a scientifically statistical atmosphere. Westergaard descends to no defense of statistical method or of the science of statistics as a whole. He expends little energy in the time-honored discussion whether statistics is a science or a method. He accepts things as they are, and having cleared the way by a characteristically painstaking historical discussion, plunges boldly into his thesis, which is, in brief, that while certain mathematical laws must be recognized and utilized, yet they should be applied with due regard to actual conditions of social development.

This, it seems to me, is Westergaard's particular contribution in the present work, even more than his clarion call to statisticians to be up and doing, or his interesting discussions of the many and varied social phenomena of which he shows so remarkable a grasp. Let me illustrate from various sections of the article.

The several laws described by Westergaard are shown each to have a place in statistical method, dovetailing with other laws in such manner as to enable statisticians to arrive at definite results and conclusions. Take the law of error, which is both based on, and limited by, the calculus of probabilities. By use of these mathematical factors, which were largely unknown to, or at least unused by, earlier statisticians, the present-day student of social movements can often test the accuracy of his facts almost as closely as can the chemist his

laboratory reactions or the biologist his developments in the embryo. Granted a sufficiently large group of observations, and a sufficiently cautious regard for all the complicated social factors involved, the statistician can plot the progress of society with what practically amounts to mathematical nicety. If the earlier statisticians failed to make adequate use of these laws, it was because they did not understand their full significance. Further, as Westergaard points out, where they did utilize the laws they failed to appreciate that their application to social phenomena required an understanding of all the conditions involved, and of the variability of such conditions. It follows that the modern statistician can appeal to a wider range of collateral sciences for assistance in adjusting his laws, such as physiology, biology, psychology, and the like.

Again, the law of interpolation frequently enables the student of statistics to fill out gaps in his data almost to the final fraction. In common with other laws, Westergaard points out that this law should not be regarded as something working wholly apart from human developments, but, as he says, "this interpolation should keep as near real life as possible, *taking everything into consideration* which might prove of importance." The italics are mine, and serve to emphasize the author's insistence, more often implied than stated, upon the absolute dependence of the statistical student on things as they are. He should not be a scholar put into his laboratory to grind out laws and methods, but a live student of human nature and human customs, adjusting his statistical theories in accordance with the *mores* of his time.

Closely related to the calculus of probabilities, in fact, almost a part of it, is the law of expectation, or actuarial law. This vital law of society enables the statistician to judge whether society is progressing or retrogressing, whether the death rate is falling and the birth rate rising, or vice versa, whether the laws of heritage are or are not operating in accordance with theory and schedule—in fact, rightly applied, it gives the social student a marvelous motion picture of society in its dynamic phases. This indicates the importance of the law, and the absolute necessity of its being accurately calculated and applied to actual conditions.

In fine, Westergaard holds that the earlier statisticians were really the predecessors of modern social students, even when they devoted themselves exclusively to the mathematics of probabilities, or to the intricacies of political arithmetic, or more particularly to the description of the comparative conditions of different states and countries. Even the last group, who were ridiculed as "Tabellenknechte," or slaves of the table, were in their feeble way working toward the same goal as the modern statistician, namely, toward an accurate picture of the condition of society. So, today, the statistician should not be a mere slave to his statistical columns, with nose buried between the covers of his logarithmic tables, but should be a wide awake and sympathetic social alchemist, applying ascertained facts to the social problems of his times. Yet Westergaard would have the statistician appreciate his limitations. As a social student he may apply the facts he has ascertained as a statistician, but the statistician's primary duty is to secure the facts. He "must be content if he can show that certain groups of numbers show marked differences, leaving it" to others to explain the differences. This does not inhibit him from utilizing, as a social scientist, the facts he has gained as a statistician.

It must be clear, from this brief analysis of Westergaard's thesis, that he has made society his lifelong study and devotion, for his volume breathes the joy and zest underlying the study of social progress by means of accurate scientific methods, and the book in unabridged form can hardly fail to be an inspiration both to statisticians and students of social development the world over. If there are criticisms to be offered of his article, the chief is his utilization of certain theoretical studies by biologists and others that would seem only remotely akin to the working of definite laws; but such defects as these seem to me fully offset by the sound optimism of his principal thesis and conclusions.

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By Carl J. West, *Ohio State University.*

It is often desired to estimate the value of a population for intercensal years. How is it to be done? Since a direct enumeration is out of the question some indirect method must be employed. As a basis for an estimate it may be assumed, for instance, that the ten year increase has proceeded by equal annual additions, or that there has been a constant annual rate of increase. Before the question of the method to be used can be definitely answered, the statistician must be familiar with the advantages and the disadvantages of each method.

A second problem of importance to the statistician making use of census data is that of distributing the excessive numbers appearing at the even ages and especially at certain ages, at age 40 for instance. No method of reënumeration can assist in the solution of this problem because the error is essentially inherent in the only practicable methods of enumeration. It might among other possible assumptions be assumed that the excess at age 40 was drawn from ages 41 and 42, or that a graphic solution is the best. Ordinarily not until quite complicated methods are made use of in removing these and other irregularities can the census figures be said to show the approximate age distribution.

It is not necessary to bring forward repeated illustrations to show that a statistician must be more than a mere enumerator, that even in the most simple statistical work some theory is absolutely necessary. It must be equally clear that instead of much of the personal and tentative methods of present day statistics there should be a consistent body of statistical theory which would be generally understood and applied. What is needed is more articles and more books and treatises presenting the theory and methods of statistics in a simple style. And further and above everything else the *theory and methods*

should be presented as a distinct science—the science of statistics. Only when statistics is generally recognized as a science worthy of equal rank with physics and chemistry and biology will it secure the proper respect and recognition without which the labor of statisticians is partially in vain.

Let us examine the article, "Scope and Method of Statistics," with the purpose of discussing Westergaard's idea of the necessity for the development of an exact science of statistics.

We begin with the theory of representative statistics. As a class economists have hesitated to make use of representative statistics, perhaps because little use can be made of such material without reference to certain very practical points of method. When can a selection be considered as representative, as typical of the whole class of individuals from which it is drawn? How is the significance of a variation in a sample or selection to be determined? are some of the questions which confront the statistician. "What remains to be done is simply to develop a theory of these representative enumerations, stating the limits of the deviations from the true proportions and showing how to approach as nearly as possible to the truth. In fact in many cases it will be practically impossible to do without representative statistics."

The matter of inaccuracies in the data is always one of great concern to the careful statistician. Note the statements of Dr. Westergaard: "a certain pessimism is often encountered among official statisticians and economists who attempt to draw conclusions from statistics. It is maintained that the inaccuracies are often so great that it is impossible to get any reliable results. Here, then, is another important problem in the theory of statistics, viz., to determine the significance of the inaccuracies, to state to what extent it is possible to draw conclusions from statistical date in spite of their imperfections. Here it would be most useful to form a *theory of the applicability of imperfect data.** For it can easily be shown that even extremely incorrect data may under certain circumstances allow of perfectly safe conclusions."

The author's discussion of the "law of error" in Section IV

*Italics are mine.

is of far reaching significance. To illustrate the nature of the discussion suppose of 400 patients with a certain disease 200 have been treated with a serum and 200 have not been so treated. Of the first class suppose further that 125 recover and of the second class only 100 patients recover. Is the observed excess of recoveries among the patients of the first class sufficient to show the efficacy of the serum treatment? In a certain community the ratio of male to female births in one year is 110 to 100 instead of the more usual ratio of 105 or so to the 100. Can one, because of this larger ratio, be certain that some cause is present which tends to increase the relative proportion of male births or is the observed excess merely a chance fluctuation, just as when tossing coins a run of heads may occur? As the basis for a fundamental theory of these variations Professor Westergaard advocates the use of the standard deviation, or "mean error," of a binomial series taken in connection with a normal law, or "law of error" as he terms it. He presents a strong case for the use of this method of measuring variations. It is of course true that not all data even approximate to the binomial law, but actual trial and experimentation show that even so the standard deviation constitutes a fairly reliable measure of variability. Altogether there appears to be a no more adaptable basis for a general theory of variability than that furnished by the standard deviation.

The idea of the author that statistical data can be broken up into a series of normal distributions is of extremely high importance as a theoretical basis for the study of a statistical distribution. The analysis of a given set of statistical measurements by identifying the component variations and values as constituent, normal distributions is one of the most significant ways imaginable for extracting information from stubborn data. It gives an arithmetic background and general basis for the work. The statistician with the theory of the normal distribution in mind feels more confident, feels that he has a standardized viewpoint, so to speak. Probably no statistical distribution can be said to have yielded all its information until it is broken up into the constituent "laws of error," or normal distributions. It may here be of interest

to define, in statistical terminology, a normal distribution. *A series of measurements is said to obey the normal law when each measurement,* X, *is equal to the sum of a relatively large number of smaller elements where each constituent element has a value chosen at random.** Thus the height of a person is the sum of the lengths of certain bones, of the widths of certain cartilages, and each of these elemental values is, within rather narrow limits, about as likely to take one value as another. Hence it would be anticipated that a distribution of height would be normal. The just mentioned value to statistics of basing the theory on a normal distribution is not lessened by the fact that in practice it is ordinarily not easy to split a distribution into its constituent normal distributions.

The author shows the true instinct of the statistician when he says at one place "On the whole we may here lay down the principle that this interpolation *first of all should keep as near real life as is possible, taking everything into consideration which might be of importance";*† at another place, "The best means of treating the changes of a variable in a series of observations is the method used by Daniell Bernoulli and Duvillard. They used the fiction that all variations were continuous so that it was possible to use the differential calculus. . . . This method is of great value as it *simplifies many problems extremely.*† . . . The main difficulty at present seems rather to consist in providing sufficiently complete and correct observations than in dealing with these materials afterwards."

The tendency of the author to systematize, to reduce to general principles, is shown by the sentence, "Still more complicated are the problems of economic statistics for here we frequently have *one dimension more,*† we have to deal not only with numbers of quantities but with their value in money as well."

Again, "Here we have many methods of procedure. *In determining which method is the best it is a good principle always to keep as near as possible to the original observations, never leaving them out of sight, with as few intermediate links as possible."*

*Edgeworth, "Law of Error," Cambridge Phil. Trans., 1905.
†Italics are mine.

On the formula of correlation we find, "The formula of correlation* will prove useful in all cases where the points are grouped nearly around a straight line; . . . *still we must not forget that this formula removes us somewhat from the original data and that it does not relieve us from the necessity of making a close investigation of these observations.* On the whole the formula of correlation does not introduce any new principle; by tabulating and grouping the observations we can easily establish as a rule the fact of correlation without the use of the formula." The author does not mention the value of a diagram—a scatter or correlation diagram, and he is apparently not familiar with the properties of the correlation ratio, but his viewpoint is fundamentally that of seeking the underlying principles of the science and of discussing each proposed method with reference to those principles.

Again with the same purposes in mind he writes with reference, apparently, to Professor Pearson's elaborate methods for determining correlation for non-measurable characteristics such as hair color, disposition, material wealth, and so on, "but at present we are more in need of statistical data than of theoretical investigations. This may sound curious in view of the immense mass of details which are published every day by the numerous statistical institutions all over the world, but to a great extent all these reports are repetitions, so to speak, of older investigations, most of them made in a single mold. . . . Of course we cannot do without these myriads of statistical volumes, they have, at least the greatest bulk of them, their local claim to exist;—but beyond these reports numerous problems are waiting to be solved and it will require much patience and much careful work in gathering the necessary materials."

It would perhaps be better to say that there is equal need of statistical data and theoretical investigations. For until the methods are clearly in mind the search for data is likely to continue to be a more or less haphazard groping in the dark. There must be some model, some plan, for the collection of data and there is nothing which can give a greater

*The coefficient of correlation alone is referred to here. C. J. W.

unity to the gathering of the material than can generally accepted and applied methods.

In conclusion it may be said that Dr. Westergaard in this article shows a thorough appreciation of the needs and difficulties of practical statistical work together with that scientific sympathy and spirit which makes all his work count towards the development of statistics into a distinct methodological science.

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By G. P. Watkins, New York, N. Y.

The special viewpoint of Westergaard would by many be called mathematical, or he would be more definitely described as one who thinks of statistics as an application of the mathematical theory of probabilities. The description is substantially correct. Its appropriateness appears in the article under discussion and would be borne out by an examination of the author's *Grundzüge der Theorie der Statistik*. But such a summary statement is seldom adequate. Some statisticians are statisticians, and some are mathematicians. Of course the implied opposition needs qualification. Indeed the most that need be claimed for it is some relative significance. The point of present interest is that Westergaard belongs with the statisticians. He brings to the aid of statistics certain mathematical conceptions, but he has no preference for the processes and expressions of algebraic mathematics and uses them only to serve his statistical purposes. This is not always true of the mathematical statistician. It is a very superficial view that supposes, because both mathematics and statistics are concerned with numbers, competence in the one involves competence in the other. Great familiarity with abstract numbers and great satisfaction in dealing with them do not promote skill in handling and interpreting numbers standing for quantities and quantitative relations of concrete things. That a statistician should be "quick at figures" and fond of equations and mathematical formulas is no more true than the other extreme view that only a physician can be competent in vital statistics, only a teacher in educational statistics, only a farmer in statistics of agriculture, etc. There is an element of truth in the second viewpoint. Concrete knowledge of the things to which statistical numbers relate is of great importance. The mistakes that are most commonly used to discredit statisticians (though they are not in fact very harmful) are due to lack of sufficient

knowledge of the concrete situations to which the numbers dealt with relate. But one can know something about farming without having spent years handling a pitchfork, and about the railroads without having been a brakeman, etc. Moreover, statistics has methods and a technique of its own —distinct from those of mathematics—with which professional familiarity is as necessary for the statistician as is the professional familiarity of the physician with disease, of the farmer with crops, etc.

The frequency and appropriateness of Westergaard's references to the necessity of keeping the original data in sight is a conspicuous characteristic of his article and the one that has suggested the foregoing remarks upon the fundamental concreteness of statistical numbers. A death rate, for example, should not be thought of as a mere ratio. Its meaning depends upon age, sex, race, etc. The "etc." is not a mere form. The statistician would like to differentiate and classify *ad infinitum* the deaths and the persons exposed to death, and find the correlation, if any, between the various pairs of classifications. The quality of statistical analysis is not secured by correctness of calculations, nor by appropriate application of mathematical theory and skillful manipulation of mathematical formulas. Accuracy of observations is more important. But statistical accuracy does not mean mathematical exactness. It means rather the reduction of biased error to a minimum, and also—but this may usually be taken for granted—the use of large enough aggregates so that unbiased errors cancel each other. The more or less selective character of observation, statistical and other, and the varied composition of concrete numbers—the two factors working together—constitute the great source of wrong statistical conclusions. The "probable error" in which the mathematical statistician is particularly interested is of comparatively little significance.

There seems to be no occasion to add anything to what Westergaard has to say of the significance of the probable error for statistical analysis. But there is one matter of terminology in relation to which I would express dissent. The reference to sample statistics as "representative statistics"

seems to me infelicitous. The adjective should refer to function and quality rather than to a situation and a numerical relation in which the quality is chiefly assumed, the only reason for alleging that samples are representative being that their selection is presumed to be unbiased. Moreover, we need the word in a truer sense for *representative numbers*, where the process by which they are obtained does have explicit reference to the function and quality in question. I refer to "averages" of all sorts, mathematical and unmathematical, weighted (or true) and unweighted (and apparently hap-hazard). Representative numbers, in the preferred sense, and relative numbers are the two great instruments of statistical analysis. Representative numbers are such as provide a condensed substitute, in some one or another use of the figures, for the complete aggregate. An average *represents* an aggregate. We try to obtain sample numbers that will represent the comprehensive aggregate where it is impossible to observe statistically the latter as a whole. But the statistical process in this case assumes rather than assures the representativeness desired.

One matter that is quite pertinent to the viewpoint of Westergaard's article seems to me not sufficiently developed. A statistical discussion of "probable error" should not stop at the application of mathematical theory. It should consider the *margin of error* in a broader sense. "Probable error" in the mathematical sense deals with accidental and unbiased and relatively small errors. But the statistician who has had to do with the interpretation of and compilation from enumeration and report schedules does not worry about such unrepresentativeness of his results as may be due to a large "probable" error. Most government statistical offices deal, not with samples, but with comprehensive or complete aggregates, which are also large enough to get the full benefit of the "law of large numbers." Where the "probable error" is negligible, however, the *margin of error* may be large. The practically important meaning of accuracy relates to the latter situation. The accuracy of a census may be tested by repeated sample enumerations, by known or determinable coloring or inefficiency in administration, by internal evidences

of duplication or omission, by the known functional relation of true results to figures obtained by other methods, etc. But all such errors are presumably biased and they tend to be too large to be left to chance cancellation. The problem is not mathematical.

I am not able to say whether Westergaard's "method of expected cases" is his original contribution to statistics or not. Perhaps the method is not in principle different enough from the "standard calculation," generally known in its application to the "correction of the death rate," to make the question of originality worth particular consideration. The "correction" method is substantially the inverse of the method of expected cases. At any rate the development and use of the latter by Westergaard affords an example that ought to influence other fields besides that of vital statistics. I obtained my conception of the importance of the method from a reading of his *Mortalität und Morbilität* (2d edition, 1901). Section 18 of chapter I introduces the idea, and the frequent apt use of the method throughout is not the least reason why the book deservedly takes the highest rank among statistical monographs. The method deserves a distinct and important place in general statistical theory. It is both simple and of broad applicability. I have found what is in effect the same method useful in studying street railway traffic increases. It should be of interest for the study of rates of growth in general. In allowing for the effect of annexed data upon per cent. increases the principle under whatever name must be taken into consideration. It should be noted that the method is not only elementary, as regards the mathematics involved, but that the required direction of attention is to the concrete.

THE FEDERAL FARM LOAN ACT.

By Walter B. Palmer, *Special Agent, Bureau of Foreign and Domestic Commerce, United States Department of Commerce.*

The Federal Farm Loan Act. The passage of the Federal Reserve act, which provided a more elastic currency system for the commercial needs of the country, has been followed within three years by the enactment of the Federal Farm Loan bill, which furnishes facilities for farmers to obtain credit on far more advantageous terms than have been offered to them heretofore. The latter bill is intended to place agriculture on an equal footing financially with other industries. It passed both branches of Congress by practically a unanimous vote. Considered from financial and economic standpoints the Federal Farm Loan act is one of the most important measures ever enacted by Congress.

Until the passage of the Federal Reserve act, in December, 1913, national banks were prohibited from making loans on real estate security, and farmers found it difficult to secure credit accommodations at reasonable rates of interest. The only security that farmers can give for a loan is usually real estate, and frequently they are charged exorbitant rates of interest, in violation of usury laws. Recent investigation by committees of Congress and the Department of Agriculture showed that the average rates of interest on farm mortgages, including commission, varied from 5.3 per cent., in New Hampshire, to 9 and even 10 per cent. in certain Southern and Western States; in some instances they were as high as 25 per cent.

The Federal Reserve act permits member banks to make loans secured by mortgages on improved farm lands, for periods of not exceeding five years, to the extent of only 25 per cent. of their capital and surplus. This makes available for such loans less than $500,000,000, whereas existing farm mortgages amount to over $2,000,000,000. Federal Reserve banks are organized principally for the benefit of the commercial and manufacturing classes. They extend or con-

tract their loans in response to the rise and fall in the supply of deposits. They must be prepared to return deposits on demand, and it is important for them to invest these deposits in only short time loans.

Federal Reserve banks have, therefore, shown little disposition to make loans on farm mortgages. Some state banks and savings institutions make such loans and permit them to run five years, but usually insist on notes payable on demand, so that the bank funds will not be tied up in case of sudden calls for the return of deposits. The farmer borrower is, therefore, in constant dread that he will be compelled to repay his loan at a time when it will cause him inconvenience or perhaps loss. When loans on farm mortgages are renewed the farmer must pay additional fees and frequently new commissions. In some states the mortgagor must pay taxes on the mortgage as well as on the land. In other states the mortgagee must pay taxes on the mortgage, which, of course, are reflected in the rate of interest.

The census of 1910 showed that the rural population of the United States then was 49,000,000 and the urban population (population of incorporated places, including New England towns, having 2,500 inhabitants) 42,000,000. The farming class, constituting about half the total population, needs loans for the purchase of additional land, the erection of farm buildings, the construction of roads, drains, and fences, and the purchase of machinery and stock. If farmers can obtain loans for such improvements on reasonable terms, agricultural operations will be greatly extended, and lands made more productive, with the result that farm products will be more plentiful. This will have far reaching effects, for with increased production consumers would be benefited by lower prices. Country life will be made more attractive and the tendency of farmers' sons to seek employment in the cities will be checked.

Of all the leading industries agriculture has been more hampered in obtaining credit for operating expenses than any other. According to the census, the value of farm products in 1909 amounted to $9,000,000,000, and of manufactured products $21,000,000,000, or over twice as much. But the

amount invested in farm lands, buildings and equipment, $41,000,000,000, was over twice the capital employed in manufacturing, $18,000,000,000.

Many farmers compelled to borrow money to carry on farming operations would have been ruined by the high rates of interest and commission charges that they have been forced to pay had it not been for the great increase in the value of farm property during recent years. The Federal census reports that during the 10 years from 1900 to 1910 the average value of farm lands in the United States increased from $15.57 to $32.40 per acre, or 108 per cent.

The percentage of mortgaged farms increased during these 10 years from 31 to 33.6 per cent., though this was partly due to the breaking up of Western ranches and Southern plantations into smaller holdings. The percentages of all mortgaged farms for which reports concerning mortgages were obtained by census enumerators are shown in the following table, these percentages being based on the combined total for farms free from mortgage and farms mortgaged in the different sections of the United States:

PER CENT. OF MORTGAGED FARMS.

Section.	1910.	1900.	1890.
United States	33.6	31.0	28.2
New England	34.9	34.1	28.2
Middle Atlantic	38.3	40.3	37.0
East North Central	40.9	39.4	37.6
West North Central	46.1	44.3	48.0
South Atlantic	18.8	16.8	7.4
East South Central	22.7	17.0	4.5
West South Central	30.6	18.2	4.8
Mountain	20.8	14.4	14.1
Pacific	36.8	27.6	28.7
The North	41.9	40.9	40.3
The South	23.5	17.2	5.7
The West	28.6	21.7	23.1
East of the Mississippi River	30.9	29.7	24.8
West of the Mississippi River	38.0	33.4	34.7

In every geographic division except the Middle Atlantic the proportion of farms mortgaged was greater in 1910 than

in 1900 and in every division except the West North Central the proportion was greater in 1910 than in 1890.

There is an abundant supply of funds which, with perfect security, could be used for investment in long term mortgage loans at low rates of interest, such as the savings of people earning wages or salaries, the proceeds of life insurance in the hands of widows and other beneficiaries, funds belonging to estates, minors, and wards in chancery, funds of insurance companies and benevolent societies, endowments of colleges, hospitals and other institutions, trust funds, and assets handled by receivers in bankruptcy. The custodians of such funds, which aggregate an enormous sum, have not usually made loans to farmers direct, but largely through investment companies charging high rates of interest and commissions.

The Federal Farm Loan act puts it in the power of those interested and those who have funds to invest to extend to farmers the credit at reasonable rates which they need as much as it is needed by merchants and manufacturers. The investment may be made in farm loan bonds, and the proceeds from the bonds used for accommodating farmers, who in the aggregate produce a great part of the wealth of the country. The bonds issued under federal supervision, protected by every safeguard, legally acceptable as security for trust funds, and exempt from all taxation, will offer a most attractive form of investment, and investors as well as farmers will be gainers.

All loans made under the Federal Farm Loan act will be made on an amortization plan, by which the principal is repaid in annual or semi annual installments during a period of from 5 to 40 years, interest payments being made at the same time. Experience teaches that a borrower is much more likely to meet his loan obligations without default if he pays the principal in small regular installments at fixed periods, instead of meeting the whole obligation after the expiration of several years.

The Federal Farm Loan Board has supervision over federal land banks, which, in turn, have supervision over national farm loan associations. Loans are made by the banks through the associations to members of the associations. The banks are authorized to issue bonds, secured by the first mortgages

they hold, and to lend to farmers giving first mortgage security the proceeds from the sale of the bonds.

The rate of interest on the mortgages shall not be more than 1 per cent. over the rate of interest of the bonds. For example, if the current rate of interest on farm loan bonds is 4 per cent., the bank may charge not to exceed 5 per cent. to farmers who borrow. This margin of 1 per cent. on the amount of unpaid principal is available to the bank for its expenses and earnings, but it will probably be reduced in practice. The act provides that the interest on bonds issued by the banks shall not exceed 5 per cent., so that the rate of interest paid by mortgagors will in no instance exceed 6 per cent., and this may be reduced by the dividends of the bank.

The national farm loan associations, which recommend loans by the federal land banks, are coöperative organizations, and are managed with a minimum of expense. Members of such associations may be expected to see that the best available men in their respective communities are selected as directors, and to keep careful watch lest unsafe loans be made with resulting personal loss to all members.

In Europe much has been accomplished by legislation that encourages coöperation. There are societies to organize cooperative institutions and lecturers who go about to spread the coöperative gospel. There are coöperative insurance companies, coöperative buying and coöperative selling organizations, coöperative markets, creameries, and breeding establishments and the like. In most European countries coöperative land mortgage and rural credit systems have been in successful operation for many years. The Federal Farm Loan act is but one factor in the vast field for coöperation on a national basis that lies before the American farmer, but it will greatly stimulate the coöperative spirit and will lead to the formation of other coöperative enterprises on a large scale.

The fact that American agriculturists were not provided with suitable credit accommodations has long been recognized, but the subject did not receive general attention until 1912. The passage of the Federal Farm Loan bill is the result of

four years of effort and publicity. Mr. David Lubin, who had been the American delegate to the International Institute of Agriculture, was invited to make an address on the subject of a rural credit system to the Southern Commercial Congress at its annual convention held at Nashville in 1912. He accepted on condition that the representatives from the various states should assemble a week prior to the convention and devote that time to the study of the subject. Twenty seven states sent representatives to the conference.

Mr. Lubin addressed the conference, explaining rural credit systems in foreign countries, and afterwards devoted much attention to the proposition of establishing such a system in the United States. The convention decided that a commission, composed of two men from each state, should go to Europe to study the subject in countries where such systems were in operation. The matter was taken up with President Taft, the Department of State, the governors of the various states and with organizations of farmers throughout the country, and the plan received general endorsement. It was endorsed in the platforms of the three major political parties in their national conventions in 1912.

Later, the agricultural bill passed by Congress contained provisions for a Federal Commission of seven members to be appointed by the President, "to coöperate with the American Commission assembled under the auspices of the Southern Commercial Congress." The American Commission, composed of two men from each of 35 states and four Canadian provinces, and five members of the Federal Commission sailed from New York in April, 1913. At Rome they were received by the King and Queen of Italy at the home of the International Institute of Agriculture. Through delegates to the institute, plans were made for a tour of 14 European countries. The commissions were aided by every facility the various governments could offer, and they returned to the United States in July, 1913.

A great mass of valuable, authoritative data was collected by the two commissions. Elaborate reports were made by the commissions—by the Federal Commission to Congress and by the American Commission to the governors of all

states,—and the data and reports were published as Senate documents, and widely distributed, especially among farmers and agricultural organizations and journals. This publicity aroused wide interest in the proposition. The measure was strongly endorsed by President Wilson, as it had been by President Taft. President Wilson's annual message to Congress, in December, 1913, was largely devoted to the subject.

The report of the Federal Commission, containing the bill recommended, was presented, in January, 1914, to the Senate by Senator Duncan U. Fletcher, of Florida, and to the House by Representative Ralph W. Moss of Indiana. Senator Fletcher was chairman of both commissions that visited Europe, and Mr. Moss a member of the Federal Commission. Senator Fletcher was President of the Southern Commercial Congress in 1912, when the movement started, and has been since then.

The agricultural appropriation bill enacted in 1915 provided for a joint committee on rural credits, to consider all bills on the subject and to report to Congress. This joint committee consisted of Senators Owen of Oklahoma, Hollis of New Hampshire, Gore of Oklahoma, Smith of Georgia, Nelson of Minnesota, and Brady of Idaho; and Representatives Glass of Virginia, Phelan of Massachusetts, Lever of South Carolina, Moss of Indiana, Hayes of California, and Hawley of Oregon. Representative Carter Glass was made chairman of this joint committee, Representative Ralph W. Moss of a sub-committee on personal rural credits and Senator Henry F. Hollis of a sub-committee on land mortgage loans. Through hearings before these committees, and by means of a wide distribution of literature on the subject, the whole country became interested as rarely before in a measure, and a realizing sense of the need for legislation became nation wide.

The draft of a revised bill and a report by the sub-committee on land mortgage loans were adopted by the joint committee on rural credits and presented to Congress in January 1916. The report, signed by Senator Hollis for the sub-committee, concluded with the following summary:

It is believed that the system of land banks outlined in the proposed bill affords a safe and attractive farm loan bond

for the investing public; low interest rates, long term mortgages, and easy payments for the farmer; low cost of administration; simplicity of organization and operation; adaptability to the needs of every section; and stimulation to the spirit of generous coöperation among farmers.

The Senate and House referred the bill to their committees on banking and currency. The bill, when reported with a few changes, was passed in the Senate by a vote of 57 to 5 and in the House by a vote of 295 to 10. The conference committee consisted of Senators Owen, Hollis, Fletcher, Nelson, and Gronna, and Representatives Glass, Phelan, Moss, Hayes, and Platt. The conference report was agreed to by viva voce vote in the Senate, and with only 12 votes recorded against it in the House. The bill was approved by President Wilson on July 17, 1916.

The Federal Farm Loan Board. A bureau of the Department of the Treasury, called the Farm Loan Bureau, will execute the provisions of the act, under the general supervision of a Federal Farm Loan Board, with headquarters in Washington. This board consists of five members, of whom one is the Secretary of the Treasury, who is chairman *ex officio*. The other four members are appointed by the President of the United States, by and with the advice and consent of the Senate, and not more than two may be appointed from one political party. One of the four is designated by the President as the Farm Loan Commissioner, who is the active executive officer of the board.

One of the four members appointed by the President serves for two years, one for four years, one for six years, and one for eight years, and thereafter each of these four members will serve for eight years, unless sooner removed by the President for cause. These four members receive a salary of $10,000 annually, and must devote their entire time to the business of the board, and not engage in any other banking or mortgage loan business.

The Federal Farm Loan Board has power:

To organize and charter federal land banks, to charter national farm loan associations and joint stock land banks, and to exercise general supervisory authority over them.

To review and alter the rates of interest to be charged for loans made by federal land banks, these rates to be as uniform as practicable.

To make rules respecting the charges to borrowers for expenses in appraisal, determination of title and recording.

To grant or refuse to federal land banks or joint stock land banks authority to make any specific issue of farm loan bonds.

Upon request of the Federal Farm Loan Board, the Secretary of the Treasury is authorized to make deposits for the temporary use of any federal land bank, at current rates of interest for such deposits, to be secured by farm loan bonds or other collateral satisfactory to him. Certificates of such deposits held by him shall be redeemed at his discretion. The aggregate of Government funds thus deposited may not exceed $6,000,000 at any one time.

The sum of $100,000, or so much thereof as may be necessary, is appropriated for carrying into effect the purposes of the act. The salaries and expenses of the members and employees of the Federal Farm Loan Board, and of a farm loan registrar for each federal land bank district, and of land bank examiners are to be paid by the United States. No funds of the United States may be used for paying any other salaries or expenses of the federal farm loan system. The officers and employees mentioned are excepted from civil service rules, but the President has the power to place them in the classified service.

It is the duty of the Federal Farm Loan Board to prepare amortization tables which are to be used by national farm loan associations, federal land banks, and joint stock land banks, to prepare forms on which such associations and banks report to it quarterly, and annually to make a full report to the Speaker of the House of Representatives, to be printed for the information of Congress.

Federal Land Banks. The Federal Farm Loan Board is directed to divide the continental United States, excluding Alaska, into 12 districts, no district containing a fractional part of any state, and to establish in each district a federal land bank, with its principal office in such city as the board

shall designate. Subject to the approval of the Federal Farm Loan Board, any federal land bank may establish branches within its district. Each district bank, before beginning business, must have a subscribed capital of not less than $750,000, divided into shares of $5 each. The capital may be increased from time to time. Stock subscriptions are required from national farm loan associations, and may be received from individuals, firms or corporations or from the government of any state, and, on organization of the bank, from the United States.

If within thirty days after the opening of books of subscription for the capital stock, any part of the $750,000 has not been subscribed, the Secretary of the Treasury shall subscribe the balance on behalf of the United States. While funds of the United States may thus be used in completing the minimum stock subscription on the organization of a federal land bank, provision is made for the retirement of the stock held by the United States when additional subscriptions are received from other sources. Government owned stock receives no dividends, but other stock shares in dividend distributions.

Each national farm loan association and the Government of the United States shall be entitled to one vote for each share of stock held by it in deciding all questions at meetings of stockholders, but no other shareholder is permitted to vote. Stock owned by the United States is to be voted by the Farm Loan Commissioner, as directed by the Federal Farm Loan Board.

Every federal land bank has power to borrow money, to give security therefor, and to pay interest thereon; to accept deposits of securities or current funds from national farm loan associations holding its stock, but to pay no interest thereon; to issue, subject to the approval of the Federal Farm Loan Board, and to sell farm loan bonds, which must be secured by first mortgages upon farm land deposited with the farm loan registrar for the district; to invest its funds in the purchase of first mortgages on farms in its district; to buy and sell United States bonds and to deposit its securities, and its current funds subject to check, with any member

bank of the Federal Reserve System, and to receive interest thereon. When necessary or convenient, a federal land bank may acquire and dispose of real or personal property, and may lease the same in part for revenue purposes, but, without the special approval of the Federal Farm Loan Board, may not retain for longer than five years possession of any real estate purchased or acquired to secure a debt.

No federal land bank has power to accept deposits of current funds payable upon demand except from its own stockholders, to make loans except on first mortgages through national farm loan associations, to issue bonds in excess of 20 times its capital or surplus, or to receive from any national farm loan association additional mortgages when the principal remaining unpaid upon mortgages already received from such association exceeds 20 times the amount of its capital stock owned by such association.

Bonds issued by federal land banks may be in denominations of $25, $50, $100, $500 and $1,000, in series of not less than $50,000. They shall have interest coupons attached, payable semi annually, and the rate of interest must not exceed 5 per cent. per annum.

All bonds are issued against collateral security of United States Government bonds, or first mortgages on farm lands, at least equal in amount to the bonds issued. The Federal Farm Loan Board may, at any time, call upon any land bank for additional security to protect the bonds it has issued. The security is made still stronger by a provision that when any farm loan bank shall default in paying the interest coupons on its bonds, the other such banks shall be liable to pay them in equitable proportion, and shall be liable also to pay proportionately the unpaid principal of farm loan bonds issued by any such bank which shall not be paid after the assets of the bank shall have been liquidated and distributed.

Not less than 25 per cent. of the capital of any land bank for which stock is outstanding in the name of national farm loan associations must be held in readily marketable securities or other liquid assets. Five per cent. of the capital must be invested in United States Government bonds.

Each federal land bank temporarily must have five directors and, when the subscriptions of national farm loan associations to its stock have reached $100,000, nine directors, of whom six, called local directors, are nominated and elected by the directors of the national farm loan associations in the district. These six local directors designate two of their number to serve one year, two to serve two years, and two to serve three years, their successors to be elected for terms of three years. The other three directors, called district directors, are appointed by the Federal Farm Loan Board and represent the public interest. In case of a tie the Farm Loan Commissioner determines the choice. The Federal Farm Loan Board designates one of the district directors to serve for three years and to act as chairman, one to serve for two years, and one for one year, their successors to be appointed for terms of three years. A director must have been a resident of the district for at least two years, and during his term of office may not engage in any other banking or mortgage loan business. At least one district director shall be experienced in farming and be actually engaged in farming in the district.

The board of directors, subject to the approval of the Federal Farm Loan Board, fix the salaries of the directors, officers, or employees, to be paid from the general funds of the federal land bank.

National Farm Loan Associations. In any farm loan district 10 or more natural persons may unite to form a corporation known as a national farm loan association, which shall be composed only of those who are owners or about to become owners of farm lands and of those who are borrowers or about to become borrowers on such land. An application to the federal land bank for a charter must be accompanied by a subscription to the capital stock of the bank equal to 5 per cent. of the aggregate sum desired on mortgage loans, which sum may not be less than $20,000.

Upon receipt of the stock subscriptions and of articles of association, the federal land bank sends an appraiser to investigate the solvency and character of the applicants for a charter, and then determines whether a charter for the

association should be granted. The recommendation of the land bank is forwarded to the Federal Farm Loan Board. If it is unfavorable, a charter is refused; if favorable, the board grants a charter, unless good cause is shown why this should not be done, and designates the territory in which the association may make loans. Upon receipt of its charter, such national farm association is empowered to receive from the federal land bank of the district sums to be loaned to its members on first mortgage farm land security.

A national farm loan association is organized on a coöperative basis. Its capital stock, consisting of shares of a par value of $5 each, may be increased from time to time for the purpose of securing additional loans for members. Each shareholder participates ratably in dividends earned and distributed, and is entitled to one vote on each share of stock held by him at all elections of directors and in deciding all questions at meetings of shareholders, but in no case to more than twenty votes. Shareholders are individually and ratably held responsible for all debts and obligations of the association to the extent of the par value of the stock owned by them, in addition to the amount paid in and represented by their shares.

Every national farm loan association has power to indorse, and thereby to become liable for the payment of first mortgages accepted from its shareholders by the federal land bank of its district to secure loans on farm lands, and to issue certificates of deposits of current funds, bearing interest for not longer than one year, at not to exceed 4 per cent. per annum, convertable into farm loan bonds when presented at the district land bank. Such deposits must forthwith be transmitted to the bank, and invested by it in first mortgages or in the purchase of bonds issued by a federal land bank.

Every national farm loan association must, out of its earnings, semi annually carry to reserve account a sum not less than 10 per cent. of the net earnings until there is a reserve of 20 per cent. of the outstanding stock issue, and thereafter 2 per cent. of the net earnings is to be added annually to this account. After the deduction of this reserve any association may declare a dividend to shareholders of the

whole or any part of the balance of its net earnings. The reserve is to be invested under rules prescribed by the Federal Farm Land Board.

Any national farm loan association may make application to the land bank of its district for loans not exceeding in the aggregate one fourth of its total stock holdings in the bank, and the bank may make the loan and charge interest at not exceeding 6 per cent. per annum.

Every national farm loan association shall elect a board of not less than five directors, who elect a president, a vice-president, a secretary-treasurer, and a loan committee of three members. All the directors and officers except the secretary-treasurer must be shareholders in the association and residents in its territory.

The directors and all officers of the association, except the secretary-treasurer, serve without compensation unless the payment of salaries to them shall be approved by the Federal Farm Land Board. The salary of the secretary treasurer, fixed by the board of directors, and the reasonable expenses of his office and of the loan committee and other officers are to be paid from the general funds of the association.

The secretary-treasurer must furnish a surety bond, and it is his duty to act as custodian of the funds of the association and to deposit the same in such bank as the board of directors may designate, to pay over to borrowers sums received for their accounts from the federal land bank, to collect, receipt for, and transmit to the federal land bank payments of interest, amortization installments, or principal arising from farm mortgage loans, and to be custodian of the stock certificates, securities, and records of the association. He must also make a report to the Federal Farm Loan Board, on specified forms, quarterly and when requested by the board. From time to time he is to assure himself that the loans made through the association are applied as set forth in the applications for loans by borrowers. He must report to the district land bank the failure of any borrower to comply with the terms of his mortgage or to pay taxes on the property mortgaged.

Method of Obtaining Loans. Loans made through national farm loan associations to its members by federal land banks,

may not be for less than $100 nor more than $10,000 to any one borrower, and not for less than 5 years nor more than 40 years. Loans in any bank district must be secured by first mortgages on farm land within the district, and must not exceed 50 per cent. of the value of the land mortgaged and 20 per cent. of the value of the permanent, insured improvements thereon, the value to be ascertained by appraisal, the basis of appraisal to be the earning power of the land for agricultural purposes. No loan may be made to any person who is not at the time, or shortly to become, engaged in the cultivation of the land mortgaged. The par value of shares of stock in national farm loan associations is $5 each, and every borrower must own one share for each $100 borrowed. In case of the sale of the mortgaged land, the federal land bank may permit the mortgage and stock interest of the vendor to be assumed by the purchaser. In case of the death of the mortgagor, his heirs or legal representatives have the option, within 60 days, of assuming the mortgage and stock interests of the deceased.

Loans may be made for the following purposes only:

To purchase land for agricultural uses.

To purchase the equipment, fertilizers, and live stock necessary for the proper and reasonable operation of the mortgaged farm, and to provide buildings for the improvement of farm lands, the terms "equipment" and "improvement" to be defined by the Federal Farm Loan Board.

To liquidate indebtedness of the owner of the land mortgaged incurred for the foregoing mentioned purposes.

Any person desiring to secure a loan must make application for membership in a national farm loan association, and state, on a prescribed form, the objects to which the proceeds of the loan are to be applied. He must also subscribe for one share of the association's capital stock for each $100, or major fractional part thereof, of the amount that he desires to borrow. The value of these shares, and the cost of appraisal, legal fees and recording charges are paid by the borrower, or, by arrangement with the federal land bank lending the money, may be included in the loan and deducted from the amount of the loan, provided that the whole amount does not exceed

$10,000. The applicant may become a member of the association by a two thirds vote of its directors. If the loan be granted he, on payment for the shares, becomes the owner of them. They are held by the association as collateral security for the payment of the loan, but the borrower receives the dividends on them while outstanding.

The application for a loan is referred to the loan committee of the association, composed of three members, who examine the land offered as security and make a detailed report. No loan may be approved by the association's directors unless the committee's report is favorable. This report is submitted to the federal land bank of the district, together with the application for the loan. The bank then refers the application and the loan committee's report to one or more of the land bank's appraisers, who investigate and make a written report upon the land offered as security, and no loan may be made unless this report is favorable.

Whenever any national farm association shall desire to secure for any member a loan on first mortgage from the federal land bank of its district, it must subscribe for capital stock of the bank to the amount of 5 per cent. of the desired loan, such subscription to be paid in cash upon the loan being granted by the bank. Such shares of stock are held by the bank as collateral security for the payment of the loan, but the association is paid any dividends accruing on such stock outstanding.

Every mortgage which a borrower makes to his association to secure his loan must be indorsed by the association to the bank that grants the loan, the association thereby making itself liable for the payment of the mortgage. Every such mortgage must be on an amortization plan, by means of a fixed number of annual or semi annual installments sufficient to cover, first, a charge on the loan, at a rate not exceeding the interest rate in the last series of farm loan bonds issued by the bank granting the loan; second, a charge for administration and profits at a rate not exceeding 1 per cent. per annum on the unpaid principal, said two rates combined constituting the interest rate on the loan; and, third, such amounts to be applied on the principal as will extinguish the debt within an

agreed period, not less than 5 nor more than 40 years. Before the first issue of bonds by any bank the rate may be determined by the bank, but at no time may the interest rate on loans exceed 6 per cent. per annum, exclusive of amortization payments. Interest on defaulted payments, however, may be at the rate of 8 per cent.

Any association is allowed to retain as a commission from each interest payment on any loan indorsed by it an amount to be determined by the Federal Farm Loan Board, not to exceed one eighth of one per cent. semi annually upon the unpaid principal of the loan, any amounts so retained to be deducted from dividends payable to the association by the federal land bank.

If the borrower expends the whole or any portion of a loan for purposes other than the specified objects for which farm loans are allowed, or if he defaults in any covenant of the mortgage, the whole of the loan may, at the option of the mortgagee, become due and payable forthwith. Taxes, liens, judgments, or assessments on the mortgaged land which are paid by the mortgagee become a part of the debt and bear simple interest at 8 per cent. per annum. Every borrower must undertake to keep insured all buildings the value of which was a factor in determining the amount of his loan. Insurance must be made payable to the mortgagee, as its interest may appear at the time of loss, and at the option of the mortgagor, subject to general regulations of the Federal Farm Loan Board, sums so received may be used for reconstruction of the buildings destroyed.

In the discretion of the federal land bank making a loan, a reappraisal of the mortgaged land may be permitted at any time, and such additional loan may be granted as such reappraisal will warrant.

After five years from the date of a loan, payments in addition to amortization payments may be made in sums of $25 or any multiple thereof for the reduction of the principal, or payment of the entire loan may be made, on any regular installment date.

The shares of stock in a federal land bank paid for by a national farm loan association when it secures a loan for

one of its members from the bank may, in the discretion of the bank's directors, and with the approval of the Federal Farm Loan Board, be paid off at par and retired, and must be paid off and retired upon full payment of the mortgage loan. In such cases the national farm loan association pays off and retires the corresponding shares of its stock which were issued to the mortgagor when the loan was secured from the bank.

Whenever it shall appear to the Federal Farm Loan Board that national farm loan associations have not been formed and are not likely to be formed in any locality, because of peculiar local conditions, the board may authorize federal land banks to make loans on farm lands through agents approved by the board. No agent may be employed, however, other than an incorporated bank, trust company, mortgage company, or savings institution chartered by the state in which it has its principal office.

Loans made by such agents are subject to the same conditions and restrictions as if they were made through national farm loan associations, and each borrower must become a shareholder in the federal land bank granting the loan to the extent of 5 per cent. of the amount he borrows.

Any agent negotiating such a loan must indorse the same and become liable for the payment thereof on default by the mortgagor. When required by the federal land bank, the agent collects and forwards to it without charge all interest and amortization payments on loans the agent indorses. The bank may allow the agent the actual expenses of appraising the land offered on security, of examining and certifying the title, and of making, executing and recording the mortgage papers, and in addition one half of 1 per cent. per annum upon the unpaid principal of the loan, such commission to be deducted from dividends payable to the borrower on his stock in the bank.

If there shall be a default under the terms of any indorsed mortgage held by a federal land bank, the national farm loan association or agent through which the mortgage was received by the bank is notified of the default, and the association or agent is required within 30 days to make good the

default, either by payment in cash of the amount unpaid, or by the substitution of an equal amount of the bank's farm loan bonds, with all unmatured coupons attached.

Joint Stock Land Banks. In some localities farmers may dislike the publicity and scrutiny which are necessary in obtaining loans through national farm loan associations from a federal farm loan bank. On account of objections to coöperative undertakings, such associations may be formed slowly in some states, with the result that farmers in a wide field could not secure credit accommodation at reasonable rates.

The act, therefore, provides for the organization of joint stock land banks, with private stock subscriptions, for the purpose of lending on farm mortgage security and issuing farm loan bonds. These banks can not engage in other business or receive deposits. The United States Government will not be a shareholder, as it may be in the organization of a federal land bank. Any number of natural persons not less than ten may organize a joint stock land bank, which must have not less than five directors. Share holders are individually and ratably held responsible for all the contracts and debts of such bank to the extent of the face value of the stock they own, in addition to the amount paid for their shares.

A joint stock land bank may be chartered by the Federal Farm Loan Board when it has capital stock to the amount of at least $250,000 subscribed, and it may begin business when one half of that amount is paid in and the balance is subject to call by the board of directors. It may not issue bonds, however, until the stock subscriptions are entirely paid, and it may not issue bonds in excess of 15 times the amount of its capital and surplus.

Joint stock land banks are restricted to a charge of 6 per cent. on farm loans, as in the case of federal farm banks, and are not allowed to charge a rate of interest exceeding by more than 1 per cent. the rate of interest on their last bond issue, and they may in no case demand or receive, under any form or pretence, any commission or charge not specifically authorized by the federal farm loan act. No joint stock bank may make loans not secured by first mortgages on farm

lands within the state in which is its principal office, or within some one state contiguous to that state.

Certain restrictions on the loans of federal farm loan banks do not apply to joint stock land banks, such as limiting the amount of loans and requiring that loans be made to farmers only and that they be used for only specific purposes. The loans of joint stock land banks must be payable in amortization installments, and in other respects so far as applicable such banks are subject to the restrictions and conditions imposed on federal farm land banks.

The Federal Farm Loan Board appoints land bank examiners, who are subject to the same requirements, responsibilities, and penalties as are applicable to national bank examiners under the Federal Reserve act. They examine federal land banks and joint stock land banks at least twice each year, and national farm loan associations whenever so directed by the board.

Every federal land bank, and every joint stock land bank must semi annually carry to reserve account 25 per cent. of its net earnings, until there is a reserve of 20 per cent. of the bank's outstanding stock issue, and thereafter 5 per cent. of the bank's earnings are to be added annually to this account. After the deduction of this reserve every federal land bank or joint stock land bank may declare dividends to its shareholders of the whole or any part of the balance of its net earnings. The reserves are to be invested under rules prescribed by the Federal Farm Loan Board.

The capital and reserve or surplus of every federal land bank and of every national farm association, and the income derived therefrom, are exempt from national, state, municipal, and local taxation. Mortgages held by federal land banks and joint stock land banks to secure loans on farm lands, and farm loan bonds issued by such banks and the income derived from these mortgages and bonds are also exempt. The shares of stock in federal land banks are exempt, but under the terms of the act the shares of joint stock land banks are not exempted, and real estate however acquired by either federal land banks or joint stock land banks is not exempted.

Farm loan bonds issued by federal land banks or joint stock land banks are a lawful investment for all fiduciary and trust funds, and may be accepted as security for all public deposits. Such bonds may be bought or sold by any member bank of the Federal Reserve System.

When so designated by the Secretary of the Treasury, federal land banks and joint stock land banks may be depositories for public money, except receipts from customs, but no Government deposits may be invested in mortgage loans or farm loan bonds.

WAGES IN AMERICAN CITIES.

By W. F. Ogburn and Esther Kelly, *Reed College, Portland, Oregon.*

Studies in the comparative statistics of cities are of importance because they place emphasis on extending the realm of knowledge about our cities and because they stimulate a healthy rivalry. Readers will appreciate the value of such studies by recalling comparative statistics with which they are familiar, such as death rates, figures of illiteracy, percentages of street area paved, and growth of population. The purpose of this paper is to direct attention to a possible ranking of cities in the average amount of wages paid. The importance of knowing wage conditions is of course very great and is emphasized particularly when we remember that we are living in an age when the fight against poverty is greater than ever before, when we are setting a minimum wage in one state after another and when many investigators are claiming that low wages are the one factor to which an increasingly large number of social ills can be traced.

In Table I is seen the ranking of 37 cities of the United States in the average union wage per hour in 10 trades. The data from which the averages and ranking were computed are for May, 1914, published in Bulletin No. 171 of the United States Bureau of Labor Statistics. It would be desirable to have a larger list of cities but the range of choice was unfortunately limited to the 41 cities from which the Bureau collected statistics. The list of 41 cities was reduced to 37 in number because data for the 10 trades were not given in 4 of the cities. These cities, however, are located in 29 different states and are the largest cities in their respective sections. They also represent typical geographical divisions and typical industrial areas, such as coal regions and textile districts.

The ranking of the cities is on the basis of average hourly union wages in the following 10 trades: bricklayers, carpenters, painters, plasterers, plumbers and gas fitters, day compositors in book and job printing, night compositors in book and job

printing, day compositors in newspaper printing, day and night linotype operators. It would have been desirable to include wages for unskilled work and salaries, but these were impossible to obtain. Indeed it was not possible to obtain data for all of the 37 cities in more than these 10 trades. For the different cities data were given for 20 to 80 trades but only 10 trades were found common to all of the 37 cities.

TABLE I.

RANKING OF 37 CITIES IN AVERAGE UNION WAGES PAID IN 10 TRADES. (a)

Rank.	Cities.	Cents per Hour.
1	Seattle	71
2	San Francisco	69
3	Portland, Ore.	67
4	Salt Lake	65
5	Denver	64
6	Chicago	64
7	St. Louis	64
8	New York	64
9	Los Angeles	62
10	Kansas City, Mo.	62
11	Pittsburgh	61
12	Boston	60
13	Newark	59
14	Washington	58
15	Cincinnati	58
16	St. Paul	57
17	Memphis	57
18	Cleveland	57
19	Omaha	56
20	Indianapolis	56
21	Detroit	56
22	Birmingham	55
23	Milwaukee	54
24	Little Rock	54
25	Louisville	54
26	New Orleans	53
27	Buffalo	53
28	Springfield, Ill.	53
29	Baltimore	53
30	Jacksonville	53
31	Providence	52
32	Minneapolis	52
33	Scranton	50
34	Philadelphia	50
35	New Haven	50
36	Atlanta	43
37	Charleston, S. C.	38

(a) The 10 trades are bricklayers, carpenters, painters, plasterers, plumbers, gas fitters, day compositors in book and job printing, night compositors in book and job printing, day compositors in newspaper printing, and day and night linotype operators.

The ranking of the cities in Table I may be considered as accurate for the 10 trades. But we are interested in knowing the ranking of these cities in all trades, as well as in a selected list of 10 trades. The question then is: How satisfactory an index is the ranking in wages in 10 trades for the

ranking in wages in all trades? How true is the sample of the whole? Since data for all trades or a very large number of trades are not given for every one of the 37 cities, how can we test the truth of the ranking in 10 trades? Although wages in a very large number are not given for every one of the 37 cities, how can we test the truth of the ranking in 10 trades? Although wages in a very large number of trades are not given for every one of the 37 cities, a very good approximation can be made by resorting to the device of breaking up the list of 37 cities into small groups. For the cities of each small group, there will be a greater number of common trades than there was for the total number of cities. A ranking may then be made for the cities in each small group on the basis of a large number of trades and comparisons made with the ranking of the same cities on the basis of the 10 trades. Upon comparison, if the relative position of any given city remains practically the same in both rankings, then the ranking on the basis of 10 trades may be accepted as a good sample, a satisfactory index of the whole. The geographical divisions were found most convenient for separating the cities into small groups and a larger number of common trades were actually found. For example in the column on the left below are listed the 6 Western cities and the average wage paid in the 10 trades; the figures in the right-hand column represent the average wages in the 35 trades that were found common to the cities of this small group.

Seattle	71	San Francisco	66
San Francisco	69	Seattle	64
Portland	67	Portland	61
Salt Lake	65	Salt Lake	59
Denver	64	Denver	57
Los Angeles	62	Los Angeles	57

The only change in the ranking is in the interchange of Seattle and San Francisco in the first place. For the Western cities, then, the ranking on the basis of 10 trades may be considered as a good index of a ranking on a large number of trades.

In Table II is found the comparison in the ranking of all of the cities by groups, in the manner presented above for the Western cities.

TABLE II.

COMPARISON OF RANKING OF CITIES IN WAGES PAID IN 10 TRADES WITH RANKING IN WAGES PAID IN A LARGER NUMBER OF TRADES.

Groups of Cities and Number of Trades in Each.		Average Hourly Union Wage Paid in 10 Trades.	Average Hourly Union Wage Paid in Large Number of Trades.	Rank on Basis of 10 Trades.	Rank on Basis of Larger Number of Trades.
North Atlantic, 20	New York	64	63	1	2
	Pittsburgh	61	55	2	4
	Boston	60	59	3	3
	Newark	59	63	4	1
	Buffalo	53	51	5	6
	Providence	52	53	6	5
	Scranton	50	49	7	9
	Philadelphia	50	50	8	7
	New Haven	50	50	9	8
South Atlantic, 13	Washington	58	57	1	1
	Baltimore	53	51	2	2
	Jacksonville	53	52	3	3
	Atlanta	43	43	4	4
	Charleston	38	34	5	5
North Central, 20	Chicago	64	62	1	1
	St. Louis	64	61	2	2
	Kansas City	62	60	3	3
	Cincinnati	58	54	4	8
	St. Paul	57	53	5	10
	Cleveland	57	55	6	4
	Omaha	56	54	7	7
	Indianapolis	56	53	8	9
	Detroit	56	54	9	6
	Milwaukee	54	50	10	12
	Springfield	53	51	11	11
	Minneapolis	52	54	12	5
South Central, 16	Memphis	57	54	1	3
	Birmingham	55	55	2	2
	Little Rock	54	56	3	1
	Louisville	54	51	4	5
	New Orleans	53	54	5	4
Western, 35	Seattle	71	64	1	2
	San Francisco	69	66	2	1
	Portland	67	61	3	3
	Salt Lake	65	59	4	4
	Denver	64	57	5	5
	Los Angeles	62	57	6	6

Columns five and six show the ranking of the cities in wages on the basis of 10 trades and on the basis of a larger number of trades. Comparing the figures in these two columns it is seen that the rank of one half of the cities is the same on the basis of 10 trades as when the wages for a larger number are used and that, although one city is shifted seven places in the ranking, the average shift is one place. It is true that the

number of trades is small in the second ranking and even greater changes would be found if the whole list of cities served rather than the small groups as the basis for the ranking. The conclusion of this somewhat detailed analysis is that on the average a city in Table I is not removed by more than two or three positions from the rank it would have received in a table constructed on the basis of a larger number of trades.

One of the most striking points to be noted in Table I is the relatively low ranking of Philadelphia as compared with New York. This difference in wages paid in two large port cities in the same economic section of the country and separated by such a short distance was so unexpected that it seemed desirable to test the results further. Accordingly the two cities were compared on the basis of hourly union wages paid in 69 common trades, with a result which showed that in New York an average wage of 54 cents an hour was paid, while in Philadelphia for the same 69 trades, an average hourly wage of 43 cents was paid. Thus, using the wage for 69 common trades, the average hourly wage in New York was found to be 11 cents higher than in Philadelphia, whereas on the basis of 10 trades the difference was found to be 14 cents. The difference in the average wages paid in the two cities is slightly lowered when a larger number of trades is considered, a result which would doubtless be found to hold with all of the cities if they were similarly tested. Therefore, judging from these added data, there is no reason to think that the relative ranking of New York and Philadelphia in Table I is not approximately correct.

It should be remembered that the data in Table I show not the actual wages paid, but rather only the ranking of the cities. If there were a very large number of union trades common to all of the cities, the actual wages paid would probably be some 10 cents an hour less than the figures given in Table I. For we found that in New York the average wage of 69 trades is 10 cents less than the average wage of 10 trades and in Philadelphia the average wage of 69 trades is 7 cents less than the average wage of 10 trades.

HIGHWAY ACCIDENTS IN NEW YORK CITY DURING 1915.

By Mills E. Case, *Secretary to the Police Department, City of New York.*

The subject of highway accidents must receive a great deal of attention in the future. At the present time, the number of persons killed or injured annually in the streets of our cities is very large, but the general public is not aware of the fact. In New York City, during 1915, the fatalities from street accidents were 2½ times the number of persons murdered in the entire city, yet it is probable that the press of the city gave 10 times as much space to murders as to accidents. Unless radical measures are taken the number of accidents will increase rapidly in all large American cities, especially in the older cities which have narrow streets. The steel skeleton type of construction has made possible skyscraper office buildings and apartment houses. The old streets are inadequate for the population of the new buildings. Some relief from street congestion is obtained through rapid transit lines, especially subways. Even the dangerous motor vehicle has done something to reduce street congestion. It takes less space on the street than the vehicles and horses which it supplants; it goes faster and so occupies the space a shorter time; it carries a larger load and so makes fewer units necessary. But each successive year a larger fleet of vehicles and a larger crowd of pedestrians must use the streets; each year the resident population increases; each year the number of visitors increases; each year more women "go to business."

In the congested residential sections of our large cities, the children must play in the streets. The old fashioned house had a yard; the tenements and apartment houses of today cover practically the whole of the lot and frequently contain children enough to fill a play-ground larger than the bare lot would be. This is a condition which is getting worse every year.

Street accidents in New York City have been recorded for several years in the precinct "Arrest and Aided Record,"

but I cannot find that any compilation of accident statistics was made prior to 1914. Police Commissioner Arthur Woods saw the importance of the matter and ordered a compilation of the statistics from the precinct records for the calendar year. He also authorized a system of special reports on highway accidents, which went into effect January 1, 1915. These reports are forwarded to the Bureau of Statistics at Headquarters, where they are compiled monthly in very elaborate tables.

Prior to August, 1915, the statistics were compiled by hand; since then, they have been compiled by means of perforated cards, with the necessary machinery for handling them. During the earlier months, it was not practicable to make the detailed studies which the mechanical method rendered possible in the later months. The tables for the year, presented herewith, are necessarily limited to the extent of the early tables. A word of caution is needed concerning the comparability of these tables with the 1914 tables (published in the Annual Report of the Police Department, 1914). Because of the supervision of the Bureau of Statistics, the 1915 statistics are more nearly complete than those for 1914. An effort was made during 1915 to obtain and include the after-mortality, of which there is no record in earlier years.

The tables must be read in the light of the following limitations upon the terms used:

Accident—In these tables no accident is included unless a person is killed or injured. The records and the reports to the Bureau of Statistics include all highway accidents known to the police. The elimination of the accidents in which no person is injured leaves a fair "sample" for administrative purposes. About 20 per cent. of the reports received are eliminated from the tabulation, because no one was injured.

Highway—Includes bridges, paths in parks; excludes elevated or subway trains, stations and stairs, ferries and other boats and piers or docks which are not a part of the highway. Some accidents which occur in the highway are excluded because independent of the place of occurrence, *e. g.*, falls on sidewalk due to intoxication, illness, or play. About 5 per cent. of all reports are excluded on this ground. The eliminating

is done at the Bureau of Statistics and the precinct officers are encouraged to forward all doubtful cases.

Fatal—The 1915 tables are corrected for after-mortality reported up to January 31, 1916. The correction began to be applied in February, 1915, hence, some January fatalities are not included, because unknown. No deaths from 1914 accidents are included in these tables.

Collision—During the first six months of 1915, this term includes cases of bicycles struck by a heavier vehicle. Since then, such cases are included with "Persons Struck By—."

Cause of Accident—There is an annoying overlapping of categories here due to the fact that the original reports indicate the causes in terms of varying generality. It is not practicable to return the reports for amendment in this respect. Rather than lose the specific information by throwing it into the general classes, the Bureau has preferred to use the awkward grouping which fits the material.

HIGHWAY ACCIDENTS IN WHICH PERSONS WERE INJURED OR KILLED IN THE STREETS, SIDEWALKS, PUBLIC HIGHWAYS, OR ON THE ROADS AND WALKS IN PUBLIC PARKS, NEW YORK CITY, 1915.

Nature of Accidents.	Accidents.			Persons Killed.					Persons Injured.				
	Total.	Fatal.	Non-Fatal.	Total.	Male.	Female.	Under 6 Years.	6 Years to 16 Years.	Total.	Male.	Female.	Under 6 Years.	6 Years to 16 Years.
Persons struck by—													
Railway trains	23	13	10	13	11	2	1	2	11	8	3	1	2
Street cars	1,771	77	1,694	77	58	19	8	10	1,730	1,227	503	277	368
Motor vehicles—													
Passenger	4,865	177	4,688	178	127	51	27	72	4,830	3,733	1,097	537	1,742
Truck or delivery	1,212	105	1,107	105	82	23	15	43	1,136	885	251	144	398
Motorcycles	377	...	377	...	...	...	...	...	414	270	144	63	114
Horsedrawn vehicles—													
Passenger	249	3	246	3	3	...	1	...	252	167	85	77	92
Truck or wagon	2,441	100	2,341	100	75	25	42	37	2,379	1,772	607	629	766
Saddle horse	16	...	16	...	...	...	...	...	16	13	3	4	3
Bicycles	352	3	349	3	2	1	...	1	357	226	131	71	158
Other vehicles	72	1	71	1	1	...	...	...	71	51	20	17	26
Falls and miscellaneous injuries received while riding on, boarding or alighting from vehicles	3,026	47	2,979	47	45	2	...	1	3,026	2,230	796	33	149
Falls, etc., while stealing rides	313	9	304	9	9	...	1	7	305	296	9	46	212
Collisions	2,428	41	2,387	42	38	4	2	5	2,982	2,489	493	46	202
Other accidents to vehicles	573	19	554	21	17	4	...	...	779	636	143	10	44
Runaways	421	13	408	13	9	4	5	1	483	431	52	23	30
Non-vehicular accidents—													
Falls	3,005	24	2,981	30	20	10	...	2	3,093	1,811	1,282	190	411
Hit by falling objects	584	1	583	1	...	1	...	1	624	436	188	48	108
Bitten by dogs	120	...	120	...	...	...	...	...	125	101	24	26	64
Miscellaneous accidents	692	16	676	16	13	3	4	2	708	584	124	59	196
Totals	22,540	649	21,891	659	510	149	106	184	23,321	17,366	5,955	2,301	5,085

ACCIDENTS BY MONTHS.

Nature of Accidents.	Total.	January.	February.	March.	April.	May.	June.	July.	August.	September.	October.	November.	December.
Persons struck by—													
Railway trains	23	1	1	1	...	2	2	2	3	3	1	3	4
Street cars	1,771	116	152	184	174	140	145	151	157	149	151	125	127
Motor vehicles—													
Passenger	4,865	272	211	289	398	432	424	501	438	582	610	431	277
Truck or delivery	1,212	100	69	99	86	92	88	137	119	104	131	91	96
Motorcycle	377	7	15	23	38	46	48	47	47	40	35	22	9
Horse-drawn—													
Passenger	249	11	11	24	27	28	11	31	28	30	26	14	8
Truck or wagon	2,441	123	136	204	247	225	225	261	242	230	242	177	129
Saddle horse	16	2	1	1	2	2	...	3	...	4	1	...	...
Bicycles	352	6	13	18	39	50	47	59	45	40	23	12	...
Other vehicles	72	...	...	6	2	6	2	22	5	11	6	7	5
Falls and miscellaneous injuries received while riding on, boarding, or alighting from vehicles	3,026	192	155	202	230	300	326	363	339	323	225	184	187
Falls, etc., while stealing rides	313	9	8	22	22	27	40	34	42	42	41	18	8
Collisions	2,428	137	161	176	189	241	285	229	201	197	209	211	192
Other accidents to vehicles	573	23	26	41	39	51	49	53	53	58	69	61	50
Runaways	421	25	22	40	40	38	46	50	33	25	34	34	34
Non-vehicular accidents—													
Falls	3,005	387	263	300	297	201	209	207	204	187	204	195	351
Hit by falling objects	584	41	33	41	64	67	73	58	61	46	31	36	33
Bitten by dogs	120	6	4	6	11	15	22	16	6	11	10	8	5
Miscellaneous accidents	602	122	78	48	51	78	76	67	48	44	35	19	26
Total	22,540	1,580	1,359	1,725	1,956	2,041	2,118	2,291	2,071	2,126	2,084	1,648	1,541

VEHICULAR ACCIDENTS.

Cause.	Number of Accidents.
Fault or incapacity of injured person—	
Stealing ride	417
Improperly riding on vehicle	203
Improperly boarding or alighting	1,290
Falling under or against vehicle	147
Running into or in front of vehicle	955
Walking, standing or sitting in street or on curb	220
Crossing street not at crossing	3,289
Skating or playing ball in street	172
Playing in street	1,099
Coasting or riding toy vehicle in street	32
Improperly riding bicycle	52
Intoxication	330
Illness	22
Crippled condition	29
Blindness	3
Deafness	26
Working in street	150
Carelessness	203
Other faults	22
Fault of driver—	
Speeding	67
Recklessness	85
On wrong side of street	175
Turning corner improperly	54
Coasting	1
Violating 8-foot law	16
Improperly towing vehicle	8
Failing to signal	14
Disregarding signal	5
Lights out	23
No flag or light on projection	1
Vehicle improperly loaded	12
On sidewalk	20
Driving into area closed to traffic	2
Obstructing traffic	1
Backing	36
Leaving auto unattended	11
Leaving horse unattended	14
Intoxication	38
Illness	2
Carelessness	105
Other faults	10
Defects of vehicles—	
Steering gear	76
Brakes	43
Wheels	47
Axle or fifth wheel	50
Harness, pole, etc	61
Unmanageable vehicle	79
Vehicle starting or stopping suddenly	75
Other defects	43
Skidding of vehicle—	
Slippery street:	
With tire chains	100
Without tire chains	227
Other causes:	
With tire chains	10
Without tire chains	80
Other causes—	
Slippery streets	26
Excavation in street	2
Obstruction in street	7
Object in street	38
Street lights out	5
Street poorly lighted	3
Other street condition	41
Fog	3
Rain	5
Snow	2
High wind	8
Traffic conditions	75
Miscellaneous causes	186
No cause given	7,486
Total	18,139

REGULARITY OF EMPLOYMENT—NEED OF STANDARDIZATION OF METHODS FOR MEASURING REGULARITY OF EMPLOYMENT.*

By N. I. Stone.

The ascertainment of annual earnings of industrial workers has been a fascinating will o' the wisp to the government statistician. The lack of information necessary for this purpose, for which the average factory pay roll is notorious, is sufficient to drive the statistician to despair. Employees are designated by arbitrary numbers which are given without distinction as to sex, occupation, method of compensation, etc. Number 25 in a garment factory may stand one week for a cutter, a highly skilled male worker, receiving $25 per week, and the following week, upon the discharge of the cutter, the same number may be given to a "finisher," a girl worker who has earned $7.88 by steady plugging at prevailing piece rates. There is nothing to indicate that number 25 of the former week and number 25 of the latter are two distinct individuals. Add to this the shifting of employees from shop to shop and it will be clear why it is practically impossible to trace the annual earnings of individual workers in most industries.

In 1913 the writer was engaged in an investigation of wages in the Dress and Waist Industry of New York City under the joint auspices of the Manufacturers' Association and the Union, with a view to standardization of piece rates paid in that industry. The pay rolls proved no exception to the general rule. The extreme fluctuations of employment, however, due to seasonal influences in that industry made some kind of an estimate of the annual earnings of the workers imperative. For this purpose the index number method was resorted to for the first time, as far as the writer is aware, in the measurement of wage fluctuations. The succeeding steps used in the process were as follows:

* Presented at the dinner of the American Statistical Association, New York City, May 5, 1916, in connection with the discussion of the subject of Standardization of Statistical Methods.

TABLE I.

FLUCTUATIONS OF EMPLOYMENT AND WAGES IN THE DRESS AND WAIST INDUSTRY FOR 1912.

Week.	Number of Employees.	Amount of Wages Paid Out.	Per Cent. (Busiest Week = 100). (a)	
			Employees.	Wages.
1	14,550	$130,484	71	53
2	15,443	153,598	75	63
3	16,190	165,549	79	67
4	16,960	180,673	83	74
5	17,849	192,382	87	78
6	18,329	202,506	89	82
7	19,146	212,972	93	87
8	19,633	221,929	96	90
9	20,183	210,614	98	98
10	20,409	244,981	99	100
11	20,521	245,494	100	100
12	20,524	245,177	100	100
13	20,374	236,951	99	97
14	19,313	194,583	94	79
15	19,298	201,271	94	82
16	19,145	208,053	93	85
17	18,974	203,595	92	83
18	18,478	192,435	90	78
19	18,078	185,635	88	76
20	17,501	178,117	85	73
21	16,761	166,905	82	68
22	16,509	156,863	80	64
23	16,350	167,384	80	68
24	16,514	169,487	80	69
25	15,961	159,534	78	65
26	14,514	141,406	71	58
27	12,292	105,559	60	43
28	10,816	95,279	53	39
29	10,304	94,149	50	38
30	11,147	105,290	54	43
31	12,429	115,379	61	47
32	13,945	138,336	68	56
33	15,677	159,322	76	65
34	16,888	177,802	82	72
35	17,686	195,967	86	80
36	18,363	191,971	89	78
37	18,107	170,519	88	69
38	18,660	197,465	91	80
39	19,030	215,996	93	88
40	19,199	215,151	94	88
41	19,388	220,809	94	90
42	19,686	216,626	96	88
43	18,695	202,148	91	82
44	17,791	188,115	87	77
45	16,769	165,736	82	68
46	16,428	167,753	80	68
47	16,378	162,302	80	66
48	15,867	149,128	77	61
49	16,364	162,455	80	66
50	16,604	169,000	81	69
51	16,879	170,462	82	69
52	16,290	150,827	79	61
Average	17,100	178,887	83.3	73

(a) In the column for employees the busiest week means the week in which the maximum number were employed; in the column for wages it means the week in which the maximum amount was paid.

First, the earnings of individual employees during the busiest week of the year were ascertained for each factory under investigation. Next, the total number of workers employed and the total amount of wages paid out each week of the year by each of those factories were ascertained. The data were arranged in tabular form, each factory forming a vertical column of 52 lines, each line representing the total wages paid by that factory during that week. The wages paid each week by the 260 shops for which these data were obtained, were then added horizontally across the table, the total weekly wages for the industry as a whole thus forming an additional vertical column.

The next step was to express the weekly wage totals for the industry in terms of index numbers. The week showing the highest wage total (the eleventh week, see Table I) was regarded as the busiest week for the industry as a whole and the wage total for that week expressed as 100. The wages for each of the remaining 51 weeks were expressed as percentages of the wages of the busiest week, the lowest week (the twenty ninth) showing only 38 per cent. and the average for the year being 73 per cent. of the wages in the busiest week of the year.

The report of the investigation was published as Bulletin 146 of the United States Bureau of Labor Statistics, and Table I is reproduced from page 160 of that Bulletin.

The annual average index number of wages, equal to 73, was used as a basis in arriving at an estimate of the average annual earnings of workers in different occupations. Thus, if the earnings of a woman operator during the busiest week were equal to, say, $10.00, her average weekly earnings during the year would be 73 per cent. of that, or $7.30, and her annual earnings 52 times $7.30, or $376.60.

In a supplementary investigation last winter, Miss Juliet Stuart Poyntz had occasion to use the raw material of the investigation covered by Bulletin 146. In order to ascertain the extent of variation of employment in different types of shops, she applied to separate shops the method used by the writer for the industry as a whole. The percentages for the individual shops calculated under her direction are shown in the following table:

TABLE II.

FLUCTUATIONS OF EMPLOYMENT IN VARIOUS SHOPS AND GROUPS OF SHOPS IN THE DRESS AND WAIST INDUSTRY.

Per Cent. of Employment.	High Grade Shops.		Low Grade Shops.		Whole Industry.	
	Number of Shops.	Per Cent. of all High Grade Shops.	Number of Shops.	Per Cent. of all Low Grade Shops.	Number of Shops.	Per Cent. of all Shops.
Less than 20	..	..	..	..	..	..
20–24	..	..	..	..	..	..
25–29	1	1.0	2	1.2	3	1.0
30–34	2	2.1	4	2.4	6	2.3
35–39	4	4.2	1	0.6	5	2.1
40–44	5	5.2	19	11.2	24	9.0
45–49	8	8.3	27	16.5	35	13.5
50–54	12	12.4	37	21.9	49	18.4
55–59	20	20.3	18	10.8	38	14.2
60–64	21	21.7	25	14.8	46	17.3
65–69	13	13.4	17	10.0	30	11.3
70–74	7	7.2	14	8.3	21	7.9
75–79	4	4.2	4	2.3	8	3.0
Total	97	100.0	168	100.0	265	100.0

Average (simple)	55.6
Average (weighted)	58.2

As will be seen from the preceding table, out of 266 shops there were only three shops with an average of 73 per cent., the figure for the industry as a whole obtained by the writer, and only four shops had higher averages, ranging from 75 to 78 per cent. The percentages for the remaining shops ranged from 27 to 72. Averaging the percentages of the separate shops and groups of shops for the industry as a whole, Miss Poyntz obtained the figure of 55.6 per cent. as a simple average and 58.2 per cent. as a weighted average.

The discrepancy between the results obtained by the two methods was so striking as to suggest at first that an error had crept into the calculations of one or the other investigation. An analysis of the tabular material, however, disclosed the interesting fact that the figures of 73 per cent. and 58.2 per cent. respectively for the industry as a whole obtained by the two methods were each mathematically correct, the discrepancy being due to the difference in method of application of the index number principle.

An examination of the diagram showing fluctuations of employment in the industry (p. 169 of Bulletin 146) will disclose the fact that the fluctuations, while roughly synchronous in different groups of shops, are not entirely coincident; the same is true as to individual shops, that is to say, while all the shops are more busy in the spring than they are in the summer, the highest peak in the employment curve may be reached during the tenth week in one shop, the ninth week in another, the eighth in a third, the eleventh in a fourth, and so on. Again, the extent of fluctuation differs for various groups of shops, and still more so for individual shops. The combining of the wages of all the shops into one total for each week had the effect of smoothing out the sharp peaks in the curves for the individual shops, the coinciding of the peak of one shop and the depression of another having a mutually neutralizing effect. In other words, the combination of what might have been 260 separate curves into one tended to flatten out the curve representing the industry as a whole. The flatter the curve, the less the fluctuations, and the nearer, of course, it approaches a straight line representing 100 per cent.

To put it in less technical terms: If the industry were so organized that workers unemployed in their own shops could be transferred to other, more busy shops, so that the entire industry formed in this respect an integral unit, the method of combining the wages of all the shops into one total for the industry as a whole would have been entirely proper; but, as a matter of fact, there is no such mobility of labor in the industry during the season. At the end of the season, a considerable number of workers do leave their places in search of better positions in other shops; but, while the season lasts, very few workers leave their shops. This means that the regularity of employment for the bulk of the workers in the industry is measured not by the quantity of work available in the industry as a whole, but by that to be had in their respective shops; and it, therefore, seems that the summary of conditions in the industry as a whole can be correctly formed by arranging the annual percentages of the separate shops, as was done by Miss Poyntz and not by the method used by the writer.

This runs counter to accepted statistical standards according to which the averaging of percentages is wrong and only absolute numbers should be dealt with.

The fact that the method adopted by the writer has since been followed in investigations of other industries conducted by the Federal Bureau of Labor Statistics as shown in its reports in Bulletins 183 and 190 lends added interest to the question of what is the correct method of measuring regularity of employment.

It is hoped that this account may lead to further discussion of the subject and contribute to the standardization of methods in the measurement of one of the most important of economic phenomena,—the fluctuation of employment and its effect upon annual earnings, in regard to which there is an extreme paucity of accurate data.

REVIEWS AND NOTES.

The Mortality from Cancer throughout the World. Frederick L. Hoffman. Newark, N. J.: The Prudential Press. Pp. 826+xv.

The statistical method of eliminating error by the force of massing large volumes of data is excellently exemplified by Frederick L. Hoffman's monumental monograph on cancer; for he includes in his survey a review of statistical returns for some 26 per cent. of the total population of the world.

Vital statistics from South American cities, from African provinces, and from Balkan states at first, of course, inspire a certain amount of suspicion. Yet the paradox enunciated by F. A. Woods,—to the effect that the more faulty the original source of data the more reliable are any conclusions to which they lead in spite of such faults,—is essentially a sound one. I assume this to mean that if the probable error or the element of random deviation is large, the causes which nevertheless produce a significant correlation must be very important indeed. This of course assumes no *constant* error, and the criticism of cancer statistics is based on the suspicion that such a constant error does exist in the shape of a tendency to better diagnosis. Many statisticians, following King and Newsholme and the British Imperial Cancer Research workers, have on this ground doubted the alleged increase in this disease, but the evidence brought forward by Mr. Hoffman certainly goes far to substantiate his conviction that the increase is a real one. The combined cancer rate for a group of European and Australasian countries increased from 44.8 per 100,000 in 1881 to 90.4 in 1911. Neither the changes in age distribution nor the improvement in medical diagnosis which have taken place can reasonably be held to account for such an enormous difference. The reported death rate is low among primitive peoples and where statistics are available its recent incidence varies from a standardized death rate of between 50 and 60 for Japan and Italy to rates between 100 and 110 for Bavaria and Switzerland. In countries of the latter type its increase has now been checked. In other words the recorded cancer rate tends to approach 100 or thereabouts. All this would in general fit in fairly well with the explanation of better diagnosis, but the detailed differences as Hoffman shows point to something more.

Particularly valuable is his analysis of cancer in different countries by the part of the body affected. Thus cancer of the stomach, liver, and oesophagus (strictly inaccessible cancers) are higher in Uruguay and Japan than in England and Wales or the United States, and more than twice as high in Switzerland, Holland, Norway, and Bavaria as in the United States, while cancer of the skin (easy of diagnosis) is much higher in the United States and England and Wales than in any of the other countries mentioned above, which facts can not be explained as due to medical diagnosis. So cancer of the female generative organs (difficult of diagnosis) is nearly as

common in Japan as in England, while cancer of the breast (easy of diagnosis) is a little more than one tenth as frequent. Both English and German statistics are cited to show that in the increase of recorded cancer during the past quarter century cancer of the accessible organs has played a more important part than that of the inaccessible organs. Altogether the impression made upon the reviewer by the mass of Mr. Hoffman's data is that the geographical and historical differences in recorded cancer death rates do for the most part represent real and significant biological facts.

The limits of a review forbid even a glance at Mr. Hoffman's discussion of the somewhat puzzling occupational statistics of cancer, or at his widely sweeping review of the geographical incidence of cancer throughout the civilized world. He finds evidence for a general increase in cancer as one proceeds northward and southward from the equator. He finds no evidence for inheritance of cancer or for any parasitical or communicable origin. The main purpose of the book is, however, to call attention to the fact "that the menace of cancer throughout the civilized world is much more serious than has generally been assumed to be the case" and to lay a foundation for future statistical study of the fundamental problems involved.

The last 500 pages of the work which constitute the appendices will prove of the greatest value to future investigators. They include the various approved classifications of tumors, the forms used for cancer records and questionnaires by leading English and American authorities, copies of popular circulars on the control of cancer, and a mass of statistical material copied from original official reports. Occupational cancer statistics are given for England and Wales, Hungary, and the United States (Prudential Insurance Co.). There are 85 pages of insurance statistics (English and American) largely by age and by the affected organ, 163 pages of cancer statistics for American states and cities (largely by organs), and 193 pages of foreign statistics (mostly crude cancer death rates by years). Finally there is a bibliography which occupies 18 closely printed pages.

Altogether this is perhaps the most impressive monograph which the science of vital statistics has yet produced in the United States. We owe much to Mr. Hoffman for his industry and enthusiasm in its preparation and to The Prudential Company for the broad vision and the public spirit which made possible the production and the publication of so notable a volume.

C.-E. A. Winslow.

Yale Medical School.

ANTHROPOMETRIC STATISTICS OF CHILDREN—AGES SIX TO FORTY-EIGHT MONTHS.

About five years ago, the *Woman's Home Companion* inaugurated a novel plan to direct public attention to the fact that our children are less perfect, physically, than is generally assumed to be the case. In its inception this movement in favor of greater emphasis on *healthy* children and less emphasis on *good-looking* children was in the form of Better Babies Contests, which were held in various localities throughout the United States, usually under the auspices of some local organization, but with the coöperation of the Better Babies Bureau of the *Woman's Home Companion.*

In these Better Babies Contests the children were examined under three groups of questions: First, with reference to mental development; second, with reference to physical defects or impairments; and, third, with reference to physical development.

The different items in these main groups were given ratings. The total or maximum rating under the mental and developmental test was 200; under physical examination for defects or impairments, the total maximum rating was 700; and under physical measurements the total maximum rating was 100—making a grand total of 1000 points.*

Early in 1914 plans were made for the holding of a Better Babies Contest in Newark, N. J., in March of that year. This contest was under the joint auspices of *Contemporary*, a prominent local organization of women, and the State Pediatric Society. The instructions, score cards, literature, medals and certificates were in this case, as in the many previous contests, provided by the *Woman's Home Companion.*

Up to this time, the table of physical measurement standards with which the children at ages under five years were compared was based upon too few numbers to warrant its use for anything more than approximate averages. This first table, also, was based upon average measurements of both sexes combined, and the difficulties of the scorers were further increased by reason of the fact that the average measurements were published only for three-month intervals for ages under one year—6 months, 9 months, and 12 months—and for four-month intervals at ages one year and upward—16 months, 20 months, 24 months, 28 months, etc.

As a result of the first Better Babies Contest held in Newark, N. J., I proposed that the table of physical standards be thoroughly revised, that it be based upon larger numbers, that males and females be considered separately, and that the average measurements be published for single months of age. These improvements were recommended as theoretically desirable even before this contest was held, and the statistics of the 542 children, ages 6 to 36 months, entered in that contest, proved that only with the aid of such a table could the physical proportions of any given child be rated fairly.

*In a later, revised score card the ratings are grouped as follows: Test I. Mental and Developmental, maximum total points 200; Test II. Measurements, maximum total points, 100; Test III. Physical Examination—maximum total, 500 points; Test IV. Oral and Dental Examination, 115 points; Test V. Eye, Ear, Nose and Throat, 85 points.

The Better Babies Bureau, of the *Woman's Home Companion*, invited me to compile such a table, and this was done in 1914, from measurements furnished by that Bureau and derived from Better Babies Contests which had been held in 23 different States during the years 1913 and 1914, including altogether 3,448 sets of measurements, of which 1,843 were of males and 1,605, of females. This first table was limited to ages 6 to 42 months, both inclusive, and as the statistics were compiled by single months the numbers for any given age were comparatively small. In spite of this fact, however, the averages gave exceptionally smooth results.

During the year 1914 a new development of the central purpose of Better Babies Contests was fostered by the Council on Public Health and Instruction, of the American Medical Association, but under the title of "Baby Health Conferences." At these conferences score cards similar to those used at the Better Babies Contests are filled out, including physical measurements. The physical measurements at both contests and conferences are made in a uniform manner and in accordance with instructions which go with the score cards. During the year 1915 these score cards of the American Medical Association were forwarded to me and the data were compiled and added to those already available through the courtesy of the *Woman's Home Companion*. In the early part of 1916 I found that the aggregate number of sets of measurements available exceeded 10,000, and this was thought to be a sufficiently large number to give satisfactory results for the purpose of compiling a revised standard anthropometric table for children of ages 6 to 48 months, both inclusive. This table, a copy of which is presented herewith, is to be published by the Committee on Public Health Education among Women, of the American Medical Association, and used in connection with the Better Babies Conferences held under the auspices of this Committee throughout the United States. This table, and the preliminary one of 1914, are thought to be the first for this country constructed in just this way—that is, from healthy children, all measured according to uniform rules and, in the main, by physicians, nurses, or other specially instructed persons. These tables differ somewhat from those referred to by pediatrists, such as Dr. Emmet L. Holt in his book entitled "Diseases of Infancy and Childhood." Dr. Holt's statistics for the first year of life are based upon observations made upon 100 healthy nursing infants at the Nursery and Child's Hospital, New York City. These children—50 males and 50 females—were weighed daily during the period of observation, and the purpose was to ascertain the normal rate of growth, which was found to be steady and fairly consistent from the third day. There were, of course, individual variations, but the average measurements —weight, height, etc.—increased gradually and consistently from the third day to the end of the first year of life.

Dr. Holt states that, "comparatively few observations have been published upon the weight during the second to the fifth years of life." He had, however, about 372 personal observations, from which he concludes that the gain in weight is about 6 lbs. during the second year, 4½ lbs. during the third year, and 4 lbs. during the fifth year. His observations were also

to the effect that throughout this period the gain in weight of girls was at about the same ratio as that of boys, but that the girls remained on the average nearly one pound lighter.

In the accompanying table it is interesting to note that during the second year of life the average increase in weight for males was 5¾ lbs., and for females, 5⅝ lbs. The increase in height during the second year of life was, for males, 4⅞ inches, and for females, 4½ inches. The accompanying table also shows that the average weight of boys at the age of six months is 1¼ lbs. more than the average weight of girls at the same age. The average height of boys at the age of six months is 26½ inches, as compared with an average height for girls at the same age of 25⅞ inches. In other words, boys at the age of six months, according to this table, are ⅝ of an inch taller on an aver-

AVERAGE PHYSICAL MEASUREMENTS. (a)
MALES.

Age, Months.	No.	Weight, Pounds.	Height.	Circumference of Head.	Chest.	Abdomen.	Diameter of Chest. Lateral.	Ant.-Post.	Length of Arm.	Leg.
6	259	18	26½	17⅞	17⅞	17⅞	5⅛	4⅛	10⅛	11⅜
7	256	19⅛	27⅛	17⅞	17⅞	17⅞	5⅛	4⅛	10⅛	11⅛
8	232	19⅛	27⅜	18	17⅞	17⅞	5⅛	4⅛	10⅛	11⅛
9	211	20⅛	28⅛	18	17⅞	17⅞	5⅛	4⅛	11	12⅛
10	209	20⅛	28⅛	18⅛	18⅛	17⅞	5⅛	4⅛	11⅛	12⅛
11	193	21⅛	29	18⅛	18⅛	17⅞	5⅛	4⅛	11⅛	12⅛
12	284	21⅛	29⅛	18⅛	18⅛	17⅞	5⅛	4⅛	11⅛	13
13	264	22⅛	29⅛	18⅛	18⅛	18⅛	5⅛	4⅛	11⅛	13⅛
14	212	23	30⅛	18⅛	18⅛	18⅛	5⅛	4⅛	12	13⅛
15	215	23⅛	30⅛	18⅛	18⅛	18⅛	6	4⅛	12⅛	13⅛
16	186	24⅛	31⅛	18⅛	18⅛	18⅛	6	4⅛	12⅛	13⅛
17	159	24⅛	31⅛	18⅛	18⅛	18⅛	6	4⅛	12⅛	14
18	181	24⅛	31⅛	19⅛	19⅛	18⅛	6	4⅛	12⅛	14⅛
19	156	25⅛	32⅛	19⅛	19⅛	18⅛	6⅛	5	12⅛	14⅛
20	170	25⅛	32⅛	19⅛	19⅛	18⅛	6⅛	5	13⅛	14⅛
21	159	25⅛	32⅛	19⅛	19⅛	18⅛	6⅛	5	13⅛	14⅛
22	160	26⅛	33⅛	19⅛	19⅛	18⅛	6⅛	5	13⅛	15
23	184	27	33⅛	19⅛	19⅛	18⅛	6⅛	5	13⅛	15⅛
24	201	27⅛	33⅛	19⅛	19⅛	19	6⅛	5⅛	13⅛	15⅛
25	150	27⅛	34	19⅛	19⅛	19⅛	6⅛	5⅛	13⅛	15⅛
26	147	28⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5⅛	13⅛	15⅛
27	126	29	34⅛	19⅛	20	19⅛	6⅛	5⅛	14⅛	15⅛
28	121	29⅛	35⅛	19⅛	20	19⅛	6⅛	5⅛	14⅛	16⅛
29	94	29⅛	35⅛	19⅛	20	19⅛	6⅛	5⅛	14⅛	16⅛
30	153	29⅛	35⅛	19⅛	20⅛	19⅛	6⅛	5⅛	14⅛	16⅛
31	113	30⅛	35⅛	19⅛	20⅛	19⅛	6⅛	5⅛	14⅛	16⅛
32	110	30⅛	36	19⅛	20⅛	19⅛	6⅛	5⅛	14⅛	16⅛
33	102	30⅛	36⅛	19⅛	20⅛	19⅛	6⅛	5⅛	14⅛	16⅛
34	113	31⅛	36⅛	19⅛	20⅛	19⅛	6⅛	5⅛	14⅛	17⅛
35	116	31⅛	36⅛	19⅛	20⅛	19⅛	6⅛	5⅛	15⅛	17⅛
36	76	32⅛	37⅛	19⅛	20⅛	19⅛	6⅛	5⅛	15⅛	17⅛
37	25	32⅛	37⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	17⅛
38	37	32⅛	37⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	17⅛
39	36	33⅛	37⅛	20	20⅛	20	6⅛	5⅛	15⅛	17⅛
40	24	33⅛	38⅛	20	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
41	21	33⅛	38⅛	20	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
42	27	33⅛	38⅛	20	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
43	17	33⅛	38⅛	20	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
44	19	34⅛	38⅛	20	21	20⅛	6⅛	5⅛	15⅛	18⅛
45	27	34⅛	39	20	21	20⅛	6⅛	5⅛	15⅛	18⅛
46	17	34⅛	39	20	21	20⅛	6⅛	5⅛	16	18⅛
47	21	35⅛	39⅛	20⅛	21	20⅛	6⅛	5⅛	16	18⅛
48	19	35⅛	39⅛	20⅛	21⅛	20⅛	6⅛	5⅛	16	18⅛

Total number of males measured, 5,602.

(a) All measurements are in inches except weight.

AVERAGE PHYSICAL MEASUREMENTS. (a)
FEMALES.

Age, Months.	No.	Weight, Pounds.	Height.	Circumference of Head.	Chest.	Abdomen.	Diameter of Chest. Lateral.	Ant.-Post.	Length of Arm.	Leg.
6	176	16¼	25⅞	16⅞	16½	16½	5¼	4¼	9⅞	11¼
7	188	17⅛	26½	17¼	17⅛	17	5¼	4¼	10¼	11⅛
8	186	18¼	27	17⅜	17⅜	17¼	5⅜	4⅜	10⅜	11⅝
9	183	19¼	27⅝	17⅜	17⅜	17⅜	5⅜	4⅜	10⅜	12
10	175	19¼	27⅞	17⅜	17½	17⅜	5⅜	4⅜	10½	12⅛
11	165	20⅛	28⅛	17⅝	17⅝	17⅜	5½	4⅜	11⅛	12⅜
12	228	20⅛	28⅝	18⅛	18	17⅜	5⅜	4⅜	11⅛	12⅝
13	208	21	29⅛	18⅛	18⅛	17⅝	5⅜	4⅜	11⅛	13
14	208	21⅛	29½	18⅛	18⅛	18	5⅜	4½	11⅛	13⅛
15	197	21⅛	30⅛	18⅛	18⅛	18	5⅜	4½	11⅝	13⅜
16	176	22⅛	30½	18⅛	18⅛	18	5⅜	4⅜	12	13⅜
17	143	22⅛	30⅞	18⅛	18⅛	18	5⅜	4⅜	12⅛	13⅝
18	178	23⅛	31⅛	18⅛	18⅜	18⅜	6	4⅜	12⅛	14
19	132	23⅛	31½	18⅜	18⅜	18⅜	6	4⅜	12⅜	14⅛
20	131	24⅛	32	18⅜	18⅝	18⅛	6	4⅜	12⅜	14⅜
21	142	24⅛	32⅛	18⅜	19	18⅜	6⅛	4⅜	12⅝	14⅜
22	125	25⅛	32⅜	18⅝	19⅛	18⅜	6⅛	4⅜	13	14⅝
23	134	25⅛	32⅞	19	19⅛	18⅝	6⅛	4⅜	13⅛	15
24	160	26⅛	33⅛	19	19⅛	19	6⅛	4⅜	13⅛	15⅛
25	126	26⅛	33⅜	19	19⅛	19	6⅛	4⅜	13⅜	15⅜
26	131	27⅛	33⅞	19⅛	19⅛	19⅛	6⅛	5	13⅜	15⅝
27	136	27⅛	33⅞	19⅛	19⅛	19⅛	6⅛	5	13⅝	15⅞
28	95	27⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5	14	16
29	95	27⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5	14	16⅛
30	132	28⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5	14	16⅛
31	109	28⅛	35⅛	19⅛	19⅛	19⅛	6⅛	5	14⅛	16⅛
32	116	29	35⅛	19⅛	19⅛	19⅛	6⅛	5	14⅛	16⅜
33	97	29⅛	35⅜	19⅛	19⅛	19⅛	6⅛	5	14⅛	16⅜
34	91	30⅛	36⅛	19⅜	20	19⅛	6⅛	5	14⅛	16⅜
35	84	30⅛	36⅛	19⅜	20	19⅛	6⅛	5	14⅛	16⅜
36	90	30⅛	36⅛	19⅜	20	19⅛	6⅛	5	14⅛	17⅛
37	28	30⅛	36⅛	19⅛	20	19⅛	6⅛	5	14⅛	17⅛
38	31	31	37	19⅛	20⅛	19⅛	6⅛	5	14⅛	17⅛
39	37	31⅛	37⅛	19⅛	20⅛	19⅛	6⅛	5⅛	15	17⅛
40	26	32	37⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	17⅛
41	16	32⅛	37⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18
42	42	32⅛	38	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
43	12	32⅛	38⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
44	22	33	38⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
45	18	33⅛	38⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
46	9	33⅛	38⅛	19⅛	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
47	12	33⅛	38⅛	19⅛	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
48	31	33⅛	39	19⅛	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛

Total number of females measured, 4,821.

(a) All measurements are in inches except weight.

age than girls. It is interesting also to note the relative changes in the circumferences of the chest and abdomen. The circumference of the abdomen in very young children generally exceeds the circumference of the chest, and this is frequently true at ages four to six months, but with increasing age up to nine months there is only a slight difference in the measurements. This is true of both sexes. From the age one year upward the excess of circumference of the chest over the circumference of the abdomen is greater for boys than for girls. For illustration, at age 24 months the average chest circumference of boys is 19⅛ inches, and the average abdomen circumference of boys is 19 inches. For girls of the same age, the chest circumference is 19⅛ inches, and the abdomen circumference is 19 inches.

The head circumference of boys was found to average approximately ½ inch more than the head circumference of girls, and this difference is persistent throughout the period of life covered by the table.

From this table the average rate of growth can be determined with approximate accuracy for any duration within the limits of the table, and the comparative rate of growth of different parts of the body is likewise ascertainable for both sexes. The great majority of the children whose measurements are included in this table were of American-born parents; the children, however, were of different stocks, including German, Irish, Swedish, some Italian, and some of various other races. It would be desirable to have similar tables for the different racial elements, and perhaps even for different sections of the country. There is some evidence that the average measurements, even of these young children, would be found to differ, in the South, for instance, as compared with the North and West, entirely aside from the probable racial differences. It would seem, however, that this table, being based upon measurements of normal healthy children in various sections of the country, should serve as a fair guide for many practical purposes without any additional refinements.

FREDERICK S. CRUM.

A PLAN FOR GATHERING STATISTICAL DATA AS A BY-PRODUCT OF ADMINISTRATIVE WORK.

We, in this country, have for a number of years been familiar with the using for scientific purposes of records kept for administrative purposes. Thus the records kept by charity organization societies, chiefly in order to increase the efficiency of their work with individual families, have been extensively used as data for the scientific study of the causes of poverty and dependency.* In some cases also items or questions, which were not necessary for administrative purposes, have been added to the record forms with the purpose in view of using the data as a basis for scientific studies of cause and effect. In such cases the supplementary data recorded for scientific uses are a sort of by-product, turned out in the regular course of administrative record keeping.

This plan of gathering data as a by-product to regular administrative work has not been so extensively used in field investigations. In at least one instance, however, effective use has been made of it—by Dr. John Robertson, Medical Officer of Health for Birmingham, England, in his study of the relation of the employment of mothers to infant mortality. The advantage of his method appears to be sufficiently great to warrant a brief description of it.

* See for example Devine's "Misery and Its Causes," pp. 230 ff., and Conyngton's "How to Help," pp. 15–16, 342–3.

In describing the plan used Doctor Robertson said in his report for 1909: "On May 10, 1907, a letter was received from the Home Office to the effect that the Home Secretary had under consideration the question of the further regulation of the industrial employment of women before and after child birth. . . . In order that fuller information as regards the effect of employment both before and after child birth on the health of the mother and child . . . might be obtained, a scheme of investigation was drafted and medical officers of health were asked to take part in the collection of data. . . . "

Having decided that two wards in his district (St. Stephen's and St. George's), because of the large number of women gainfully employed, furnished an excellent area for such an investigation, a plan or method of collecting the data had to be devised. The method chosen was rather unique, although Doctor Robertson does not appear to have appreciated this fact. Instead of employing investigators to make special visits to the homes of mothers for the particular purpose of obtaining the information needed, it was decided to simply have the health visitors, who in the course of their regular work would be visiting the mothers in the two wards anyway, ask the necessary questions and make the needed observations as a supplement to their regular duties. Thus the gathering of the data as to the relation of the employment of mothers to infant mortality was a sort of by-product of the regular health work.

The visits were made by Dr. Jesse Duncan and "two experienced health visitors." "Every baby born (with a few exceptions) was visited at "frequent intervals" and at the end of twelve months weighed. Describing the results for 1908 Doctor Duncan says: "The children born in the district are visited as soon as convenient. . . . At this visit directions are given [as usual] as to the care of the child, but [in addition] information is obtained regarding the mother's employment, previous history, husband's wages," the mother's age, the method of feeding, etc. This plan of investigation was also followed in 1909 and 1910.*

From the standpoint of accuracy and thoroughness, the advantages of this method of gathering data are clear. Doctor Robertson shows this when he says: "I desire to add that the work which Dr. Jesse Duncan [assisted by the two health visitors] has done, and upon which this report is based is characterized by its accuracy and thoroughness. She has spent her whole time among those residing in St. Stephen's and St. George's Wards, and is familiar with every phase of life in these districts. This has enabled her to check and correct the statements made by the mothers she has interviewed." This is one advantage. In another place he points out another—the fact that "close contact was maintained with each of these mothers during a whole year, and much information not available at the first visit was obtainable at subsequent visits."

* Health Department of the City of Birmingham, England: "Report on Industrial Employment of Married Women and Infant Mortality," supplement to the annual report of the department for 1909, and "Report on Infant Mortality in St. Stephen's and St. George's Wards," supplement to the annual reports of the department for 1910 and 1911.

The expense involved in gathering data by this method is probably also considerably less than that involved in the usual house to house method. Since the data gathered is a sort of by-product of the regular work of the health visitors and since the visits to the homes would be made and probably most of the necessary questions asked any way, even if no records were kept of the answers, it would appear that the additional expense involved would be chiefly that of recording and tabulating the data and preparing and printing the report.

This method of gathering data might well be adopted by many American cities in studying infant mortality and similar problems.

H. H. Hibbs, Jr.

Nashville, Tenn.

India. Moral and Material Progress and Condition of India During the Year 1913–1914. India Office. London: Eyre and Spottiswoode, July, 1915. Pp. 137.

This report is compiled from detailed accounts from each of the presidencies and districts and is presented to parliament annually. The statement is divided into 18 chapters, which summarize administrative and financial matters, sources of revenue, condition of various industries, and other aspects of life in India.

The remarkable growth of coöperative credit societies is outlined. These were primarily organized to help the cultivators and to protect the poorer classes from greedy and merciless money lenders. The membership increased from 28,629 in 1905–6 to 744,226 in 1913–14, and the working capital from £32,000 to £5,145,000. These societies help to develop not only rural communities, but urban as well. The societies, through small loans, enable artisans to purchase necessary tools and implements. This fact is of great importance since the manufacturing population has increased sufficiently to necessitate special labor legislation. In a good many respects this legislation is not much, if at all, behind that of some of our states. Factory inspection and condition of women workers are definitely covered. Child labor laws call for working papers (issued upon medical examination); a minimum age for employment (9 years); and maximum number of working hours (6 hours in textile industry, 7 in others). These facts speak not only of industrial development, but of intellectual and social growth as well. This is further emphasized by the chapter on education which shows that colleges and institutions of even higher learning can boast not only of native male, but also female students.

Defects in the vital statistics are frankly admitted. In addition to the usual causes of under registration met with in other countries,—indifference to and neglect of legislation—the Government in India is confronted with utter ignorance on the part of information gathering agencies as well as on the part of the population as a whole. Even with defective registration, the records indicate a very high birth rate (39.4 per 1,000). It is due,

however, not so much to fertility as to a large proportion of married women in the population. The death rate is also very high (28.7 per 1,000). The nature of the chief causes of death indicates that the high rate is due principally to unsanitary habits and conditions. The following table presents the principal facts of deaths by diseases:

MORTALITY FROM PRINCIPAL CAUSES OF DEATH IN INDIA, 1911-1913.

Diseases.	1913.		1912.		1911.	
	Number of Deaths.	Rate per 1,000.	Number of Deaths.	Rate per 1,000.	Number of Deaths.	Rate per 1,000.
Fevers (a)	3,983,112	16.71	3,936,085	16.49	4,207,356	17.63
Cholera	294,815	1.24	407,769	1.71	354,005	1.48
Dysentery and Diarrhoea	246,578	1.03	292,216	1.22	253,636	1.06
Respiratory Diseases	237,229	1.00	247,736	1.04	223,822	.94
Plague	198,456	.83	263,037	1.10	733,582	3.07
Smallpox	98,155	.41	89,357	.37	53,538	.25

(a) Malarial fever mostly.

The existing state of affairs, bad as it is, indicates a great improvement, thanks to the growing number of hospitals, dispensaries, Pasteur institutes, and other public welfare institutions.

Another vital factor in the life of the people is the constant internal migration of population. It is of such proportions that it affects the scale of wages which, according to the report, keeps a fairly even pace with the rise of prices. Wages are further influenced by an increasing demand for labor due to industrial development.

The report shows definite progress,—steady if slow. The Government office is to be commended for its care and perseverance in gathering the facts under conditions which are most unfavorable for obtaining reliable information, both through limited facilities and ignorance of informers.

JOSEPH RAYWID.

Switzerland-Zeitschrift fur Schweizerische Statistik, 50. u. 51 Jahrgang. Bern: Stampfli and Company. 1914 and 1915.

The fiftieth volume of this journal of Swiss statistics contains Dr. Karl Bader's article on "Groups in Statistics and in Sociology," a philosophy of classification. Heinrich Schlosser of the Museum of Industrial Arts, Zurich, discusses the application of the elements of design to the construction of graphic charts. In a review of graphic statistics at the Swiss Exposition, Bern, 1914, Dr. Hans Schorer defines his ideas of graphic illustration of statistical data. Both of these articles should interest American statisticians because of current activities in the standardization of graphic statistical practice on the part of American learned societies.

The fifty-first volume gives a review of official statistics at the 1914 Swiss Exposition by Dr. C. Muhlemann. The same article discusses the

nature and purposes of official statistics as adjuncts to government, legislation and the political sciences in general. Dr. Furlan of Basel treats the subject of reorganization of the Swiss statistical service in his article and recommends the formation of a Central Statistical Commission for the republic.

Vital statistics for the year 1914 are also given. The European War profoundly affected the marriage rate; the effect upon birth and death rates will not become statistically apparent until the data for 1915 are published.

An article published by the Swiss Statistical Bureau shows the results of the 1910 population census for the several main classes of the population with respect to country of origin, and sex and age.

The effect of the war upon Swiss industries may be seen in the statistics of manufactures and of exports in the tabular appendix to the 1915 Journal. Furlan, in another study, presents his investigation of the density of the Swiss population.

E. W. Kopf.

New York City.

1920 CENSUS IN ITS RELATION TO VITAL STATISTICS.

Dr. John W. Trask, Chairman of the Vital Statistics Section, American Public Health Association, has appointed the following committee to take under consideration the subject of vital statistics of the 1920 United States Census: Dr. W. H. Guilfoy, Chairman, Dr. Cressy L. Wilbur, Dr. John S. Fulton, Prof. Jas. W. Glover, Mr. Edwin W. Kopf, Prof. Walter F. Willcox, Mr. Frederick L. Hoffman, Mr. Franz Schneider, Jr., Dr. W. H. Frost, and Dr. Wm. H. Davis.

Among the problems which the committee will be called upon to consider carefully are:

Life Tables for the triennium, 1919–1921.

Differential fecundity and vitality of race and nativity classes of the population.

Revision of mortality statistics of the decade 1910–1919 in the light of 1920 Census results, including review of achievements in reducing mortality from preventable disease.

Mortality in occupations, triennium 1919–1921.

Service of vital statistics in social reform; sickness, mortality and other vital statistics as a foundation for social insurance.

Vital statistics of institutions for care of the sick, triennium, 1919–1921. (Hospitals, asylums for the insane, indigent, etc.)

Plans for the improvement and extension of the registration area for births and deaths.

Improvement of municipal and state vital statistics.

Revision of International Nomenclature and Classification of Diseases, applicable to the years 1919 *et seq.*

Enumeration of sicknesses on Census Day, 1920, in selected areas.

Sickness and accident statistics of insurance societies.

The necessity of the census authorities furnishing some of the large cities of the country with age groupings of the population according to nativities and other related facts.

The age groupings according to selected areas in these cities.

The age groupings according to occupations.

The appointment of the committee recalls the symposium on vital statistics for the Twelfth Census, 1900, which appeared in the March, 1897, Quarterly of this Association and contributed by Messrs. Wilbur, Willcox, Fouse, Hoffman, Abbott, and King.

E. W. KOPF.

New York City.

AMERICAN PUBLIC HEALTH ASSOCIATION.

PRELIMINARY PROGRAM OF THE SECTION ON VITAL STATISTICS, OCTOBER 24-27.

Chairman, DR. J. W. TRASK,
Washington, D. C.

Secretary, DR. LOUIS I. DUBLIN,
New York, N. Y.

FIRST SESSION.

Symposium on Enforcement of Registration of Births and Deaths and the Use of the Resulting Records.

1. The Enforcement of Registration of Births and Deaths in Metropolitan Areas—Methods, Results and Checks. Wm. H. Guilfoy.
2. The Enforcement of Registration of Births and Deaths by a State Registration Office—Methods, Results and Checks. Cressy L. Wilbur.
3. The Use of Registration Records of Births and Deaths in the Commercial, Scientific, and Legal Life of Metropolitan Populations. S. W. Wynne.
4. The Use of Mortality and Morbidity Statistics in Epidemiologic Studies. Allan W. Freeman.
5. The New United States Life Tables—Their Purpose and What They Show. Jas. W. Glover.
6. The Enforcement of Registration of Births in Indiana. J. N. Hurty.

SECOND SESSION.

Symposium on the Notification of Disease and the Utilization by the Health Department of the Information thus Obtained.

1. The Uses Made of Morbidity Reports by the Minnesota State Department of Health. A. J. Chesley.
2. The Uses Made of Morbidity Reports by the New York State Department of Health. F. M. Meader.
3. The Notification of Disease and the Uses Made of the Information thus Obtained in Municipal Public Health Administration. W. H. Price.
4. The Notification of Tuberculosis and the Supervision of Reported Cases in Massachusetts. Eugene R. Kelley.
5. The Establishment of "An Area of Known Disease Prevalence." Is it practicable and desirable at the present time? What diseases should be reported and what standards of enforcement of notification required? John S. Fulton.
6. The Registration of Mental Disorders. A. J. Rosanoff.

THIRD SESSION.

(Joint Session with Sociological Section.)

Symposium and Discussion on Methods of Collection and Expression of Data in Health and Social Investigations.

1. The Necessity for Health Standards. Dr. Donald B. Armstrong, Director, Department of Social Welfare, New York Association for Improving the Condition of the Poor.
2. The Limitation and Necessities of Socio-Health Investigations. Mr. Franz Schneider, Sanitarian, Department of Surveys and Exhibits, Russell Sage Foundation, New York City.
3. The Application of Statistical Methods to the Field of Socio-Health Research. Dr. Louis I. Dublin, Statistician, Metropolitan Life Insurance Co., New York.
4. Discussion: Led by J. W. Schereschewsky, Surgeon, United States Public Health Service; Dr. Ernest C. Meyer, International Health Commission, New York City; Dr. Haven Emerson, Commissioner of Health, New York City; Prof. C.-E. A. Winslow, Yale University, New Haven, Conn.; Prof. E. B. Phelps, Hygienic Laboratory, Washington, D. C.; Dr. W. C. Woodward, Commissioner of Health, Washington, D. C. Other discussants to be announced later.

FOURTH SESSION.

1. Reports of Committees.
2. Vital Statistics and Publicity as Exemplified by New York City's Recent Experience with Poliomyelitis. Chas. F. Bolduan.
3. The Massachusetts Endemic Index. H. R. Brown.
4. The Seasonal Distribution of Communicable Diseases Based on Morbidity Experience. John S. Fulton.
5. A Statistical Study of Diphtheria. F. S. Crum.

All communications with regard to this section should be addressed to Dr. Louis I. Dublin, 1 Madison Avenue, New York City.

Publications of the American Statistical Association.

The New Series of the Association, issued quarterly, will be sent to members of the Association on payment of the annual dues of two dollars. The subscription price is also two dollars per annum. Separate copies can also be purchased upon remittance of the advertised prices.

The following list of titles includes only the most important papers published in the several issues. There is in addition a considerable amount of miscellaneous material.

VOL. I. 1888-1889. Nos. 1-8. Pp. 492. PRICE $4.00.

No. 1. Statistics of Water Power Employed in Manufacturing in the United States. GEORGE F. SWAIN. 75 cents.

Nos. 2, 3. Park Areas and Open-Air Spaces in American Cities. E. R. L. GOULD.

Key to the Publications of the United States Census, 1790-1887; with Occasional References to Other Statistical Works. EDWARD CLARK LUNT. $1.00.

No. 4. Life Insurance in the United States. WALTER C. WRIGHT. $1.00.

No. 5. Notes on the Statistical Determination of the Causes of Poverty. AMOS G. WARNER. REMARKS. SAMUEL W. DIKE.

Statistics of Divorce in the United States and Europe. SAMUEL W. DIKE. 50 cents.

No. 6. American Railroad Statistics. ARTHUR T. HADLEY.

Statistics of Municipal Finance. HENRY B. GARDNER. 75 cents.

No. 7. Prison Statistics of the United States for 1888. ROLAND P. FALKNER. 50 cents.

No. 8. Finance Statistics of the American Commonwealths. EDWIN R. A. SELIGMAN.

Divorce in France. BENJAMIN F. KELLER.

Relief of the Poor in Germany. A. G. WARNER. $1.00.

VOL. II. 1890-1891. Nos. 9-16. Pp. 470. PRICE $4.00.

No. 9. Mortgage Statistics. GEORGE K. HOLMES. 50 cents.

No. 10. The Study of Statistics in Italian Universities. CARROLL D. WRIGHT.

Statistics of Private Corporations. ROLAND P. FALKNER. 50 cents.

Nos. 11, 12. Statistics of the Colored Race in the United States. FRANCIS A. WALKER.

How to Make Statistics Popular. CHARLES F. PIDGIN.

Retail Prices in Boston and Vicinity. WALTER F. COOK. $1.00.

No. 13. The Growth of Cities in Massachusetts. HORACE G. WADLIN.

Rate of Natural Increase of Population in United States. HERMAN HOLLERITH.

The First Census of Massachusetts. SAMUEL A. GREEN.

The Commercial Death Rate. ALBERT C. STEVENS. $1.00.

No. 14. A Basis for Statistics of Cost of Production. CARROLL D. WRIGHT.

A Formula for Predicting the Population of the United States. H. S. PRITCHETT.

Weight and Longevity. T. B. MACAULAY.

Some Results of Sanitary Legislation in England since 1875. GARY N. CALKINS. 75 cents.

No. 15. The Eleventh Census. R. P. PORTER.
Criminal Statistics. ROLAND P. FALKNER. 75 cents.

No. 16. The Census Enumeration in Prussia. CARL C. PLEHN.
A Plea for the Average. GEORGE K. HOLMES. 75 cents.

VOL. III. 1892-1893. Nos. 17-24. Pp. 614. PRICE $4.00.

No. 17. Statistics of Crime in Massachusetts. FRED G. PETTIGROVE.
Development of Statistics of Religions. M. FOURNIER DE FLAIX. TRANSLATED BY ALICE R. JACKSON.
Net Profits of Manufacturing Industries in the State of Massachusetts. FREDERICK B. HAWLEY.
Classification of Trade Statistics. FREDERICK C. HICKS. 75 cents.

Nos. 18, 19. A Statistical Inquiry concerning Domestic Service. LUCY M. SALMON.
The Theory and Practice of Price Statistics. R. P. FALKNER.
Measures of Distribution. GEORGE K. HOLMES.
Statistics of Suicides in New England. DAVIS R. DEWEY.
Semi-Annual Census of Paupers in Minnesota. H. H. HART.
75 cents.

No. 20. The Establishment of an Imperial Commission for Labor Statistics in Germany. CARL C. PLEHN.
French Statistical Publications. CHARLES D. HAZEN.
Sketch of Life of Robert W. Wood, M. D. JOHN WARD DEAN.
75 cents.

Nos. 21, 22. Observations on the Measure of Change. C. H. COOLEY.
Cost Statistics of Public Electric Lighting. VICTOR ROSEWATER.
Immigration and the Foreign Born Population. RICHMOND MAYO-SMITH.
The Value of Percentile Grades. LUTHER GULICK. $1.00.

No. 23. Murders in Massachusetts. WALDO M. COOK.
Classification of Occupations in the Census. TRANSLATION OF REPORT MADE BY M. BERTILLON.
Growth of Cities in the United States during the Decade 1880-90. CARL BOYD.
The Vital Statistics of an Apache Indian Community. 75 cents.

No. 24. Statistical Data for the Study of the Assimilation of Races and Nationalities in the United States. RICHMOND MAYO-SMITH.
Report of an International Mortality Standard, or Mortality Index. JOSEPH KÖRÖSI.
Character and Volume of the Money of the United States. MAURICE L. MUHLEMAN.
Fluctuations in the Secured Circulation of the National Banks and their Relations to the Prices and Investment Values of Bonds. CHARLES A. CONANT.
Currency Reform in Austria-Hungary. DR. JULIUS MANDELLO.
Comparative Statistics of Primary Education. E. LAVASSEUR.
Results of Recent Investigations on Prices in the United States. F. W. TAUSSIG.
Geographical Concentration: an Historic Feature of American Agriculture. JOHN HYDE.
The Course of Wages in the United States since 1840. CARROLL D. WRIGHT.
Some Recent Results in Railway Statistics in the United States. HENRY C. ADAMS.
The Condition and Needs of Statistics of Marriage and Divorce. SAMUEL W. DIKE.

Railway Statistics as Applicable to Earnings of Passenger Trains M. RIEBENACK.
Comparability of Trade Statistics of Various Countries. A. E. BATEMAN.
The Geographical Distribution of the Population of the United States. HENRY GANNETT.
A Preliminary Report on Anthropometry in the United States. EDWARD MUSSEY HARTWELL.
Remarks on the Theory of Anthropometry. FRANZ BOAZ.
On the Application to Individual School Children of the Mean Values Derived from Anthropological Measurements by the Generalizing Method. W. TOWNSEND PORTER.
Anthropometric Statistics of Amherst College. EDWARD HITCHCOCK.
An Anthropometrical Study of the Effects of Gymnastic Training on American Women. CLAËS J. ENEBUSKE.
Railway Freight Traffic Statistics. C. P. LELAND. $1.50.

VOL. IV. 1894-1895. Nos. 25-32. Pp. 387. Price $4.00.

Nos. 25, 26. The Marriage Rate in Michigan. 1870-90. WALTER F. WILCOX.
The Classification of Occupations for Census Purposes. HENRY GANNETT.
Nativity and Occupation of Members of the Massachusetts Legislature. F. H. HOWLAND. 75 cents.
No. 27. Causes of Poverty Further Considered. AMOS G. WARNER.
Congress of Hygiene and Demography. R. P. F. 50 cents.
Nos. 28, 29. Computation of the Value of the Wealth in Existence. GEORGE K. HOLMES.
Effects of Free Surrender and Loan Privileges in Life Insurance. MILES MENANDER DAWSON.
Was the Count of Population in 1890 Reasonably Correct? HENRY GANNETT. 75 cents.
No. 30. Receipts and Expenditures of Certain Wage-Earners in the Garment Trades. ISABEL EATON.
The Negro in the West Indies. FREDERICK L. HOFFMAN. 75 cents.
No. 31. Almshouse Women. MARY ROBERTS SMITH. 75 cents.
No. 32. Accidents in Factories and Elsewhere. KATHARINE PEARSON WOODS.
The Marriage Rate in Massachusetts. F. S. CRUM. 75 cents

VOL. V. 1896-1897. Nos. 33-40. Pp. 389. PRICE $4.00.

No. 33. Real Estate Values in Boston. HENRY WHITMORE.
Ethnic Influences in Vital Statistics. W. Z. RIPLEY. 75 cents.
No. 34. Railway Competition: A Problem in Statistics. H. T. NEWCOMB.
Mortality in Twenty-Three Massachusetts Cities. CHARLES E. BURNAP.
Note on Methods of Estimating Population. C. L. WILBUR. 50 cents.
Nos. 35, 36. Round Numbers in Wages and Prices. EDWARD D. JONES. $1.00.
No. 37. Remarks of President Walker at Washington.
Vital Statistics for the Twelfth United States Census. CRESSY L. WILBUR.
The Vital Statistics of the Census. WILLIAM A. KING.
General Francis A. Walker. 75 cents.

No. 38. Francis Amasa Walker. ADDRESS BY CARROLL D. WRIGHT.
Bibliography of the Writings and Reported Addresses of Francis A. Walker. 50 cents.

No. 39. Some Characteristics of Farm and Home Proprietors. GEORGE K. HOLMES.
A Trial Bibliography of the Writings of Johann Peter Süssmilch, 1707–1767. W. F. WILLCOX AND F. S. CRUM. 50 cents.

No. 40. Farm Ownership and Tenancy in the United States. EDWARD ATKINSON AND L. G. POWERS.
Vital Statistics at the American Public Health Association. 50 cents.

VOL. VI. 1898-1899. Nos. 41-48. Pp. 398. PRICE $4.00.

No. 41. Obstacles to Accurate Statistics. JAMES H. BLODGETT.
Uniformity in Census Returns. JOSEPH KÖRÖSI.
Cost of Production of Corn and Oats in Illinois in 1896. NATHAN A. WESTON. 50 cents.

No. 42. Concentration in Pig-Iron and Coal Production. E. W. PARKER.
The Annual Statistics of Manufactures in Massachusetts. HORACE G. WADLIN. 50 cents.

No. 43. Comparative Statistics of Railroad Rates. HENRY H. SWAIN
Comparative Statistics of Railroad Service under Different Kinds of Control. C. E. PREVEY. 50 cents.

No. 44. Considerations in Gathering Forestry Statistics. B. E. FERNOW. 50 cents.

No. 45. Contributions to the Study of Wage Statistics. CHARLES J. BULLOCK.
A Year of State Deficits. WORTHINGTON C. FORD. 50 cents.

No. 46. The Growth of the Population of Boston. FREDERICK A. BUSHEE.
Wage Statistics in Theory and Practice. ROLAND P. FALKNER.
A Comparative Study of the Statistics of Agriculture of the Tenth and Eleventh Census. N. I. STONE. 75 cents.

No. 47. Notes on Map Making and Graphic Representation. W. Z. RIPLEY.
The Portuguese Population in the United States. FREDERICK L. HOFFMAN.
Report on Uniform Financial School Reports. 50 cents.

No. 48. Development of the Plan for a Census of the World. JOHN HOWARD DYNES. 50 cents.

VOL. VII. 1900-1901. Nos. 49-56. Pp. 466. PRICE $4.00.

Nos. 49, 50. Statistics of College and Non-College Women. MARY ROBERTS SMITH.
The Comparative Accuracy of Different Forms of Quinquennial Age Groups. ALLYN A. YOUNG.
Aids in the Use of Government Publications. LUCIUS PAGE LANE. 75 cents.

No. 51. The Registration Laws in the Colonies of Massachusetts Bay and New Plymouth. ROBERT RENÉ KUCZYNSKI.
Tramps and Wayfarers. 50 cents.

No. 52. Advertising in the United States. SIDNEY A. SHERMAN.
The Practical Use of Vital Statistics. FREDERICK L. HOFFMAN. 75 cents.